

情報が分散共有された組織の進化的学習と  
知識創造のシミュレーション研究

課題番号 12680449

平成 12 年度～13 年度科学研究費補助金（基盤研究 (C) (2)）研究成果報告書

平成 14 年 3 月

研究代表者 高橋 真吾  
(早稲田大学理工学部)(助教授)

はしがき

本報告書は科学研究費補助金による「情報が分散共有された組織の進化的学習と知識創造のシミュレーション研究」の研究成果のうち学会誌、学会等において 12, 13 年度に発表印刷された結果をまとめたものである。

本研究の目的は、組織における情報の分散共有の状況をハイパーゲームの枠組みで記述し、そのゲームを繰り返すことで各構成員の持つ解釈枠組を「進化的に」学習していく過程を遺伝的アルゴリズムにより分析し、組織における知識創造との関係を解明することである。

本研究では、解釈枠組みの学習モデルの本質は「進化概念」にあると考え、進化的学習のための基本枠組みを提示する。それは以下のようなステップである。

- ①状況の情報を分散共有した各組織構成員が自律的に意思決定し、状況に対し行動する。
- ②行動の結果を組織内の他の構成員とコミュニケーションにより情報交換する。
- ③コミュニケーションの結果に基づいて、状況に関する知識を創造し、各解釈枠組を改善する。
- ④改善された新しい解釈枠組に基づいて各構成員の自律的意思決定を行う。

本研究で開発したモデルの特徴と研究経過は概略以下のとおりである。

まず、本研究までに開発してきたネットワーク型動的ハイパーゲームに基づいて、各ゲーム形式（利得行列，戦略集合）に関する学習過程を扱うためのモデルを複数定式化した。各モデルにおいて、認知の学習過程を遺伝的アルゴリズムによりシミュレーション分析した。

また、組織学習の基本的問題とシステムの適応過程の問題との本質関係が明らかとなったことから、さらにエージェント指向の考え方を取り入れて、ハイパーゲーム型状況を本質的に包含する一般的な分配モデルに基づく組織のモデルを定式化した。この新たな一般モデルを使って、エージェントの学習過程を限定された条件の下で遺伝的アルゴリズムを用いて分析した。

以上の研究の結果次のような成果が得られた。

1) 組織学習論で提唱されている2種類のループ学習を表現するレベルの違いを明確に区別することがモデル記述のレベルで可能となった。

2) さらにこれまでのハイパーゲームに依拠したモデルを拡張，一般化し，分配モデルに基づく組織の基本モデルを新たに定式化することで，組織における2種類の学習レベルを明示的にモデルの中で表現することができるようになった。

3) 組織に関するエージェント指向モデリングの立場からの研究と完全に接続することが可能となった。

4) 本研究による基本モデルが，組織学習の全体モデルの中で，基本的学習モジュールを記述する位置にいたることが明らかとなった。

#### 研究組織

研究代表者：高橋 真吾（早稲田大学理工学部助教授）

#### 交付決定額

平成 12 年度 1,900 千円

平成 13 年度 1,700 千円

総計 3,600 千円

#### 研究発表

##### 学会誌等

- 1) Adaptive Learning of Hypergame Situations by Using Genetic Algorithm, (U.S.Putro,K.Kijima and S.Takahashi), IEEE Trans. of Systems,Man and Cybernetics –Part A:Systems and Humans, 30(5),pp.562-572,2000.
- 2) Simulation of Adaptation Process in Hypergame Situation by Genetic Algorithm,(U.S.Putro, K.Kijima and S.Takahashi),Systems Analysis Modeling Simulation, 40(1), pp.15-37,2000.
- 3) エージェントベースモデリングのための数理的システム理論の課題,(高橋真吾),社会・経済システム, no.19, pp.46-54, 2000.
- 4) システム理論からの組織論へのアプローチー組織学習の進化的システムモデルの枠組ー (高橋真吾), 組織科学, 第 34 巻, 第 2 号, pp.59-68, 2000.
- 5) Evolutionary Learning in Agent-based Modeling, (S.Takahashi), Discrete Event Modeling and Simulation Technologies(H.S.Sarjoughian and R.E.Celler Ed.), pp.297-314, 2001.

##### 口頭発表

- 1) 組織のダブルループ学習のための学習レベルを考慮した分配モデルによる基本的考察 (高橋真吾), 平成 13 年 3 月, 『社会組織のマルチエージェントシステム分析～数理とシミュレーションからのアプローチ～』システム工学部会・知能工学部会共済研究会資料 01PG0004, pp. 121-126
- 2) 進化的アプローチを構成する遺伝オペレータの組織システムへの解釈 (高橋真吾) 平成 13 年 6 月, 経営情報学会 2001 年度春季全国研究発表大会予稿集, pp. 349-352
- 3) 相互認識を考慮した一般モデルによる組織学習レベルの考察(高橋真吾)平成 13 年 11 月, 計測自動制御学会システム・情報部門学術講演会講演論文集, pp.121-126

# Correspondence

## Adaptive Learning of Hypergame Situations Using a Genetic Algorithm

Utomo Sarjono Putro, Kyoichi Kijima, and Shingo Takahashi

**Abstract**—In this paper, we propose and examine adaptive learning procedures for supporting a group of decision makers with a common set of strategies and preferences who face uncertain behaviors of “nature.” First, we describe the decision situation as a hypergame situation, where each decision maker is explicitly assumed to have misperceptions about the nature’s set of strategies and preferences. Then, we propose three learning procedures about the nature, each of which consists of several activities. One of the activities is to choose “rational” actions based on current perceptions and rationality adopted by the decision makers, while the other activities are represented by the elements of a genetic algorithm (GA) to improve current perceptions. The three learning procedures are different from each other with respect to at least one of such activities as fitness evaluation, modified crossover, and action choice, though they use the same definition for the other GA elements. Finally, we point out that examining the simulation results how to employ preference- and strategy-oriented information is critical to obtaining good performance in clarifying the nature’s set of strategies and the outcomes most preferred by the nature.

**Index Terms**—Adaptation process, genetic algorithm, hypergame, learning.

### I. INTRODUCTION

The purpose of this paper is to propose and examine several adaptive learning procedures for supporting a group of decision makers who have a common set of strategies and preferences and face nature’s uncertain behavior. We describe the situation as a hypergame situation consisting of two parties, called a group of decision makers and the nature, respectively.

Game theory is probably the only theoretical framework that can deal with decision situations with multiple conflicting parties [6]. Traditional game theory assumes that all the parties know that they see the same game. In other words, in the theory, decision situations are modeled objectively and the main concern is with rational decision behavior.

However, it is quite usual in real situations that the decision makers perceive the nature’s actions differently and subjectively. This kind of decision situation is sometimes referred to as a soft problematic situation in the fields of systems science and operational research [4], [12]. Hypergame [13] is a typical framework to deal with soft problematic situations. Contrary to the traditional game theory, it assumes that each decision maker, though involved in the same decision situation, may see it differently and subjectively.

In this paper, we adopt the hypergame framework in order to explicitly describe each decision maker’s subjective perception about the

nature’s game (or behavior). We describe the perception as a subjective game. The decision situation we will consider in this paper is as follows. Two decision parties participate in it: one party is a group of  $n$  decision makers and the other is the nature, a blackbox to the decision makers. The decision makers of the former have incomplete information about the nature’s set of strategies and preferences, though they have a common set of strategies and preferences.

In general, there are two approaches to such decision-making situations [7]. The first is to *analyze* decision situations by capturing them from a bird’s viewpoint. In this approach, researchers are assumed to know exactly about any private information of the decision makers and the nature. For example, Intelligent Poly-Agent Learning Model (I-PALM) [6] is a typical framework taking this approach, i.e., it is proposed to analyze a dynamic process through which two parties with diversified and conflicting preferences rewrite their internal models by learning about the decision situations. The main interests of I-PALM are in misperception about some decision situation and equilibria that may emerge differently from the traditional game [10], [11], rather than in learning about the opponent’s set of strategies and preferences.

The other is to *support* a particular decision maker or a particular group of decision makers in the situation to learn about his/her/their learning. Researchers taking this viewpoint are only supplied with restricted and incomplete information, which is the same as the information that the decision maker or the group of decision makers can obtain about the decision situation.

This paper takes the second approach and tries to propose learning procedures to support the group of decision makers to improve their understanding about the nature. We assume interactions between the group of decision makers and the nature occur repetitively so that they give the decision makers opportunity to improve their present perception of the nature’s behavior.

Learning procedures we will propose for the group of decision makers to learn the nature’s game have the following features. First, they are to guide the group of decision makers to learn not only the nature’s preference over the outcomes, but also the set of strategies available for the nature.

Second, they consist of several activities. One of the activities is to choose actions based on current perceptions and rationality adopted by decision makers. The other activities are to revise perceptions about nature, which are represented by the elements of a genetic algorithm (GA) [1], [5], [9]. Since it is impossible for each decision maker to collect all relevant information at one time, he/she is only able to clarify the nature’s behavior by improving his/her perception about the behavior by trial and error. The GA is suitable because it provides a method for reorganizing knowledge according to past experience in order to improve performance.

Accordingly, the main contribution of our research is that it represents decision situations by hypergames and employs the GA as a tool for the decision makers to learn the nature’s game.

So far, we have proposed learning procedures only about the nature’s preference [14]. One of the originalities of this paper lies in that it tries to learn both the strategies available for the nature and the preference of the nature.

After formulating hypergame situation we consider in Section II, we will explain idea of adaptive learning model in Section III. Section IV, the main part of this paper, introduces three kinds of learning procedures and describes them in detail. Section V describes the design of

Manuscript received December 2, 1998; revised June 1, 2000. This work was supported in part by Grant-in-Aid for International Scientific Research 09044025 of the Ministry of Education, Japan. This paper was recommended by Associate Editor J. Oommen.

U. S. Putro and K. Kijima are with the Department of Value and Decision Science, Tokyo Institute of Technology, Tokyo 152-8552, Japan (e-mail: utomo@valdes.titech.ac.jp; kijima@valdes.titech.ac.jp).

S. Takahashi is with the Department of Management and Systems Engineering, Chiba Institute of Technology, Chiba 275-0016, Japan (e-mail: takahashi@pf.it-chiba.ac.jp).

Publisher Item Identifier S 1083-4427(00)07052-1.

the experiments and then the results and findings. In this section, we set several scenarios of the hypergame situation in order to test the learning procedures and to compare their performance with each other by simulating them in all the scenarios. Section VI gives the conclusion.

## II. HYPERGAME SITUATION

Let us assume the decision situation consisting of two parties. One party is a group of decision makers denoted by  $A = \{1, 2, \dots, n\}$ , each of which has the same set of strategies and preferences. The other party is called nature and is denoted by  $B$ . The conflicting situation then happens between each of  $A$  and  $B$  while all the members of  $A$  cooperate with each other.

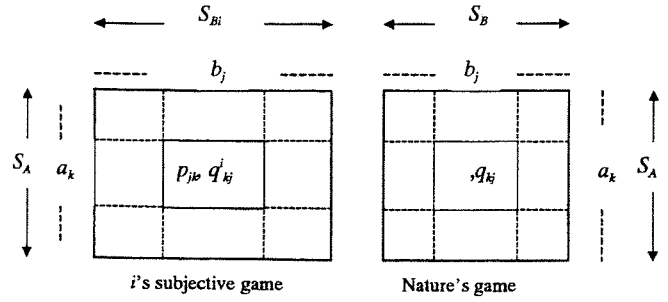


Fig. 1. Hypergame model of  $i \in A$  and nature.

### A. Assumptions of the Hypergame Situation

We assume the following for each decision maker  $i \in A$ .

- $i$  is associated with a subjective payoff matrix of the decision situation involving  $i$  and the nature. The subjective game of  $i$  can be seen in Fig. 1.
- Although  $i$  may not be sure of the nature's set of strategies  $S_B$ ,  $i$  knows it should be a subset of a certain universal set of strategies  $S$ . That is,  $i$  may be still unsure of which elements of  $S$  contained in  $S_B$ .
- $i$  perceives  $S_B$  subjectively. Hence, we will sometimes write  $S_{Bi}$  rather than  $S_B$  to emphasize it is  $i$ 's subjective perception.  $i$  knows that  $S_{Bi}$  is not an empty set.
- $i$  has the same set of strategies  $S_i$ , i.e.,  $S_i = S_A$ .
- $i$  has common preference ordering over outcomes in  $S_A \times \{b_j\}$  for each  $b_j \in S$ . For each  $b_j \in S$ ,  $i$  employs a common ordinal utility function  $P_{b_j}^A: S_A \times \{b_j\} \rightarrow \{1, 2, \dots, |S_A|\}$ , where  $P_{b_j}^A$  is bijective. We write for each  $a_k \in S_A$ ,  $P_{b_j}^A(a_k, b_j) = p_{jk}$ .
- $i$  believes that the nature has preference ordering on  $\{a_k\} \times S_{Bi}$  for each  $a_k \in S_A$ . That is,  $i$ 's perception of the nature's preference ordering is represented by an ordinal utility function  $P_{a_k}^{Bi}$  for each  $a_k \in S_A$  and each  $b_j \in S_{Bi}$  defined by  $P_{a_k}^{Bi}: \{a_k\} \times S_{Bi} \rightarrow \{1, 2, \dots, |S_{Bi}|\}$ , where  $P_{a_k}^{Bi}$  is bijective. We write  $P_{a_k}^{Bi}(a_k, b_j) = q_{kj}$ .

Consequently, the decision makers may apply various and different actions (or strategies) simultaneously, since they incorporate different subjective payoff matrices. The conflicting situation between each decision maker and nature is described by Fig. 1. We assume the response from the nature to the actions is most likely determined by the most frequent action of  $A$  at the preceding time. Formally, we define how the nature determines its current response as follows.

For each  $a_k \in S_A$ , let  $\mathcal{F}(a_k)$  represent the number of decision makers in  $A$  who applied  $a_k$  at the preceding iteration. Let us denote  $\mathcal{K}(a_k) = (\mathcal{F}(a_k)/n)$  where  $n$  is the number of decision makers in  $A$ . Then the probability of an element  $a_k \in S_A$  to be responded by nature at the current iteration is  $\mathcal{K}(a_k)$  for each  $a_k \in S_A$ . Therefore, nature selects the most preferred response from  $S_B$  over the selected element as its current response.

Accordingly, at each iteration, the nature's response lags one iteration behind the decision makers' actions; i.e., the nature is assumed to know the decision makers' common set of actions. Also, we assume that the nature's response is not influenced by the decision makers' common preference; i.e., it is not contained in the nature's game (as seen in Fig. 1).

### B. Example of the Hypergame Situation

As an illustration of a hypergame situation described above, we provide a case of a group of fishermen who want to catch a kind of fish in the sea. In this case, they are uncertain about the fishing location.

There are three actions in their common set  $S_A$  of actions:

$\leftarrow S_{B_i} \rightarrow$

		$b_1$	$b_2$	$b_3$
$S_A$ $\uparrow$ $\downarrow$	$a_1$	1, $q'_{11}$	3, $q'_{12}$	1, $q'_{13}$
	$a_2$	3, $q'_{21}$	1, $q'_{22}$	2, $q'_{23}$
	$a_3$	2, $q'_{31}$	2, $q'_{32}$	3, $q'_{33}$

Fig. 2. Fisherman  $i$ 's subjective game.

- $a_1$ : choose a fishing location randomly;
- $a_2$ : choose the last fishing location where he got a big catch;
- $a_3$ : depending on the season, choose one of the fishing locations where he got a big result.

Let the fishermen know that the nature does not apply a response beyond the universal set  $S$  of strategies containing the following elements:

- $b_1$ : a large amount of the fishes settle at a fixed location all time;
- $b_2$ : a large amount of the fishes move to another location chosen randomly after some of them have been caught by a fisherman;
- $b_3$ : a large amount of the fishes settle at one of certain locations depending on the season;
- $b_4$ : a large amount of the fishes settle at one of certain locations, and they move to another one of the certain locations after some of them have been caught by a fisherman.

At one time, a fisherman  $i$  may have a subjective payoff matrix that represents his/her perception of the nature's game, as seen in Fig. 2.

In Fig. 2,  $i$  believes that the nature's set of strategies is  $\{b_1, b_2, b_3\}$ , which is a subset of the universal set of strategies  $S$ . In other words, at the time he/she believes that it is impossible for the nature to give response  $b_4$ . By learning,  $i$  will eventually know the true set of strategies and preference of the nature, so that he/she may determine an optimal action.

## III. ADAPTIVE LEARNING MODEL

In the hypergame situation, the decision makers learn the nature's set of strategies and preferences by iterating their interactions with it. A model of learning we investigate in this paper is illustrated in Fig. 3.

For each decision maker  $i \in A$ , let us denote  $i$ 's perception of the nature's behavior for the  $t$ th iteration by  $\alpha_i^t$ . Then,  $n$ -tuple  $\alpha^t = (\alpha_1^t, \alpha_2^t, \dots, \alpha_n^t)$  represents a population of perceptions for the  $t$ th iteration. Each perception represents a decision maker's subjective game.

At each iteration, each decision maker using his/her subjective game plays a game with nature; so that for each iteration there are  $n$  interactions between the group and nature. Each decision maker chooses an action based on action choice in a learning procedure.

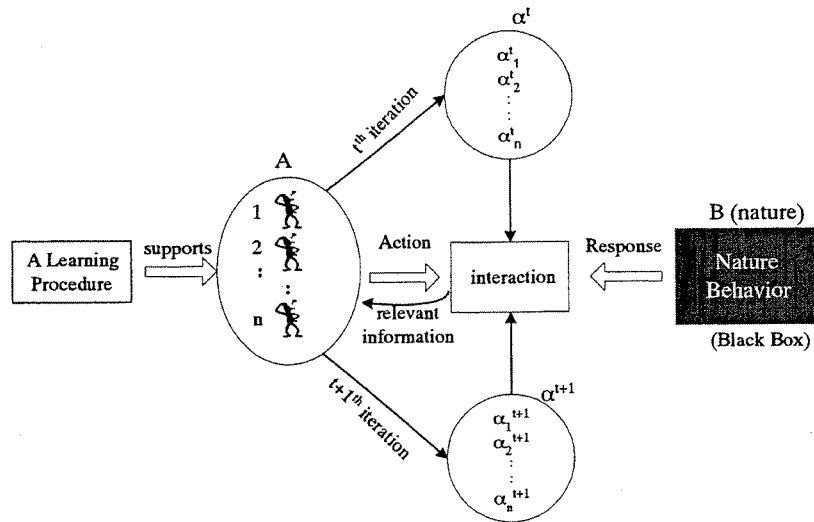


Fig. 3. Adaptive learning model.

From the interactions at the iteration, the group of decision makers may also get some important information necessary for learning. Considering the information, the group evaluates and revises the current population of perceptions from  $\alpha^t$  to  $\alpha^{t+1}$  by using the GA, based on the new perceptions the decision makers repeat their interactions with nature for the  $t + 1$ th iteration.

Accordingly, a learning procedure implies how the decision makers choose actions based on current perceptions, what kind of information is considered and shared among decision makers in  $A$ , and also represents how decision makers in  $A$  use the information to evaluate and revise the current perceptions.

By repeating the interactions, the group may or may not have chances to improve the population of perceptions close to the true nature's game. The degree of the improvement heavily depends on a learning procedure the group adopts to learn about the nature.

This paper assumes each decision maker  $i$  shares some private information with any other decision maker in  $A$ , and they agree to be guided by a procedure for effective learning. What kind of learning procedures provide the decision maker with more chances is one of the main problems handled by this study.

#### IV. LEARNING PROCEDURES

The learning procedures proposed in this paper adopt the GA to survive and evolve good perceptions of the nature's game. The GA "mimics" the natural selection process to evolve good perceptions in a population of perceptions [8].

The natural selection is a mechanism that survives the fittest individual in some population where each individual competes each other for something such as food or mating. There are unsuccessful and successful individuals as the result of the competition. The successful individuals have more chances to survive and to grow than the unsuccessful individuals. Consequently, they have a higher probability to be chosen as parents in the breeding process than the unsuccessful individuals.

During the breeding process, reproduction, crossover, and mutation may occur. Reproduction simply copies chromosomes of parents to children. Crossover and mutation make revisions to the chromosomes of parents before being passed on to children with probabilities  $p_c$  and  $p_m$ , respectively. Crossover combines chromosomes of a pair of parents, while mutation makes small local changes. Generally,  $p_m$  is much smaller than  $p_c$ . Some children have new characteristics produced by

crossover and mutation, while the other children are with simply characteristics copied from parents. The breeding process may produce children better than their parents. By repeating the competition again and again, the population will be dominated by the better individuals.

The GA is one of methods to simulate such a natural selection process [8]. The GA consists of elements such as encoding, generation of the initial population, fitness evaluation, breaking criteria, reproduction, crossover, and mutation.

Our research adopts the GA to represent how the group of decision makers revise their population of perceptions based on previous experiences.

In Darwinian evolution, the only mechanism you have is random change, which either is or is not adaptive. If it increases fitness, the change survives; otherwise, it does not. In organizations, such random adaptive processes may occur occasionally, but we can also make a conscious decision that certain things contribute to survival and other things do not and, thus, deliberately affect modification of the system [3].

The random changes are necessary because we have incomplete information about environment of the organizations. On the other hand, the conscious changes are also important in the extent that decision makers believe that they know about some of characteristics of the environment based on admissible information so far.

Accordingly, we propose an adaptation process of the group of decision makers also containing both of two kinds of changes, namely, random and conscious changes. The conscious changes need stronger revision operators to modify perception as soon as there is additional information about the nature's behavior. Accordingly, we do not use mutation in this paper, because mutation provides only small local revision and is not enough to accommodate the conscious changes in it.

We accommodate the conscious changes in our nonstandard crossover; that is to accommodate the process of how the group modifies their perceptions as soon as there is additional information. In this paper, we modify the standard crossover as follows.

- The current interaction outcome may give important information about the situations most preferred by the nature; so that it is necessary to accommodate the information to revise the current perceptions. Since a decision maker's action is based on his/her current perception, there may be some different results of interaction between the group of decision makers and nature. Accordingly, the decision makers may revise their current perceptions differently.

- By observing the variations of the nature's responses so far, a decision maker may know that current perception of the nature's set of strategies should be wrong, then crossover in some learning procedure accommodates how the decision maker revises the perception according to the information.
- Since combination process and the other revisions in crossover may produce chromosomes that imply nonordinal perceptions of the nature's preference, crossover should be also able to restore the perception of the nature's preference to the ordinal utility function defined in Section II.

In this paper, we propose three learning procedures, called LP1, LP2, and LP3, each of which consists several activities. One of the activities is to choose actions based on the perceptions and rationality adopted by the decision makers. For the action choice, we propose two kinds of rationality, that is, subjective Nash equilibrium action and random action. Action choice of a decision maker represents how he/she chooses an action based on his/her current perception of the nature's game.

The other activities are to update perceptions, which are represented by GA elements. That is, in this paper individuals in the natural selection are represented by perceptions (or subjective games of decision makers).

For our purpose, the elements of the GA are interpreted as follows. Encoding is to represent a decision maker's perception about the nature's preference and set of strategies by a bit string or a chromosome. How decision maker  $i$  calculates the fitness of his/her perception into an element in  $[0, 1]$  is implied by fitness evaluation. Selection in our research is a mechanism by which decision makers survive good perceptions in the population of perceptions. And finally, by crossover a pair of decision makers in a group combine their perceptions and modify the perceptions based on experiences so far.

The activities of each learning procedure are organized as follows.

- 1) *Stage 1—Initialization*: Generate randomly an initial population of  $n$  perceptions, each of which is associated with a decision maker's subjective game. Each perception is encoded by a bit string (or a chromosome).
- 2) *Stage 2—Action choice*: Given his/her current perception, each decision maker  $i$  in the group plays a game with nature. The way the decision maker selects an action is defined by a learning behavior adopted by the group.
- 3) *Stage 3—Update*:

- a) Revise the current population of perceptions by the GA as following procedure.
  - i) *Fitness evaluation*: Calculate fitness score of each perception in the current population by a fitness function depending on which learning behavior adopted by the group.
  - ii) *Breeding process*: Repeat the procedure below until producing  $n$  children (or  $n$  new perceptions produced by the revisions of current perceptions).

*Selection*: Select two perceptions from the current population as parents, in such a way that the fitter perceptions have the higher chances to be chosen as parents, and a perception may be chosen more than once.

*Reproduction or crossover*: Generate random number  $u$  from Uniform  $(0, 1)$  distribution. If  $u < p_c$ , do crossover: the mating of the two parents to produce two new perceptions; where  $p_c$  is crossover probability. Otherwise,

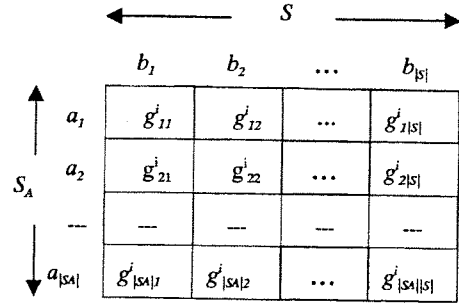


Fig. 4. Matrix form of chromosome of  $i$ .

$$\underbrace{g_{11}^i g_{21}^i \dots g_{(|S_A|-1)1}^i g_{|S_A|1}^i}_{\text{genes in the 1}^{st} \text{ column}} \quad \underbrace{g_{12}^i g_{22}^i \dots g_{(|S_A|-1)2}^i g_{|S_A|2}^i}_{\text{genes in the 2}^{nd} \text{ column}} \quad \dots \quad \underbrace{g_{1|S|}^i g_{2|S|}^i \dots g_{(|S_A|-1)|S|}^i g_{|S_A||S|}^i}_{\text{genes in the } |S|^{th} \text{ column}}$$

Fig. 5. Chromosome of a decision maker  $i$ .

do reproduction: two children are produced by simply copying the two parents.

- b) Replace the current population of perceptions with the population of  $n$  children.

LP1, LP2, and LP3 are different from each other with respect to at least one of such elements as fitness evaluation, modified crossover, and action choice, though they use the same definition for the other GA elements. Now, we will describe activities of the learning procedures employed.

#### A. Encoding

We define the chromosome so that it is able to be decoded to give information about the decision maker's perception of the nature's set of strategies and preferences in his/her subjective game.

We define  $i$ 's chromosome as follows. With each  $(a_m, b_k) \in S_A \times S$ , we associate a function  $g_i$  ( $i = 1, 2, \dots, n$ ) defined by  $g_i: \{(a_m, b_k) | (a_m, b_k) \in S_A \times S\} \rightarrow \{0, 1, 2, \dots, |S_{Bi}|\}$ , which is such that for each  $a_m \in S_A$

$$g_i(a_m, b_k) = \begin{cases} P_{a_m}^{Bi}(a_m, b_k) & \text{if } b_k \in S_{Bi} \\ 0 & \text{otherwise.} \end{cases}$$

We call  $g_i(a_m, b_k)$  a gene for each  $(a_m, b_k) \in S_A \times S$ . Accordingly, a zero gene represents a situation that does not exist in  $i$ 's subjective game; and a nonzero gene encodes a situation that exists in  $i$ 's subjective game, and it also represents  $i$ 's perception about the nature's preference on the situation. Therefore, each gene is associated with a situation  $(a_m, b_k)$  that possibly exists in  $i$ 's subjective game, that is illustrated by the matrix in Fig. 4. Each gene occupies a cell in the matrix.  $g_{mk}^i$  in the matrix denotes a gene  $g_i(a_m, b_k)$ , where  $a_m \in S_A$  is located in the  $m$ th line and  $b_k \in S$  is located in the  $k$ th column of the matrix.

Then we define chromosome of decision maker  $i$  called  $C_i$  consisting of  $|S_A \times S|$  genes as described in Fig. 5.

$C_i$  is able to be decoded to give information about  $i$ 's perception of the nature's set of strategies and preferences in his/her subjective payoff matrix. For example, if the first  $|S_A|$  genes in  $C_i$  (or the genes located in the first column of the matrix in Fig. 4) are zeros, then  $C_i$  gives information that  $b_1$  is not in  $S_{Bi}$ ; accordingly, all the outcomes  $S_A \times \{b_1\}$  are not in  $i$ 's subjective payoff matrix. And if the second  $|S_A|$  genes (or the genes located in the second column of the matrix) are nonzeros, then  $b_2$  is in  $S_{Bi}$ ; accordingly, the outcomes  $S_A \times \{b_2\}$  are in his/her subjective payoff matrix. The nonzero genes in  $C_i$  also represent  $i$ 's perception of the nature's preference over outcomes in his/her subjective payoff matrix.

We use the GA to evolve good perceptions (encoded by chromosomes), namely, the perceptions that correctly perceive the nature's set of strategies and preferences. We call  $S_{Bi}$  implied by  $C_i$  a decision maker  $i$ 's correct perception of  $S_B$  if  $S_{Bi} = S_B$ . A decision maker  $i$  correctly perceives an outcome  $(a_m, b_k)$  most preferred by nature if for each  $b_l \in S_B$ ,  $P_{a_m}^B(a_m, b_k) \geq P_{a_m}^B(a_m, b_l)$  and for each  $b_j \in S_{Bi}$ ,  $P_{a_m}^{Bi}(a_m, b_k) \geq P_{a_m}^{Bi}(a_m, b_j)$ .

### B. Generation of the Initial Population

Initial population represents the first samples of the space of all the possible perceptions about the nature's game. Since at the initial iteration each decision maker has no information about the nature's game, they generate the  $n$  initial perceptions encoded by chromosomes at random.

### C. Action Choices

An action choice of a decision maker  $i$  is based on his/her current perception of the nature's game and rationality adopted by him/her. This paper uses two kinds of rationality, called subjective Nash equilibrium action and random action, respectively.

$(a_t^*, b_q^*) \in S_A \times S_{Bi}$  is called a subjective Nash equilibrium in  $i$ 's game if  $(\forall a_k \in S_A)(P_{b_q^*}^A(a_t^*, b_q^*) \geq P_{b_q^*}^A(a_k, b_q^*))$  and  $(\forall b_j \in S_{Bi})(P_{a_t^*}^{Bi}(a_t^*, b_q^*) \geq P_{a_t^*}^{Bi}(a_t^*, b_j))$ .

We denote a set of all the subjective Nash equilibria in  $i$ 's game by  $N_i$ .  $a_t^* \in S_A$  is called a subjective Nash equilibrium action if there is  $b_q^* \in S_{Bi}$  such that  $(a_t^*, b_q^*) \in N_i$ .

On the other hand,  $a_t \in S_A$  is called a random action if  $a_t$  is selected randomly.

By the definition of preference ordering of both of the decision makers and the nature in Section II, the subjective Nash equilibria in  $i$ 's subjective game represent outcomes on which  $i$  believes that he/she and the nature have the highest preference.

Action choice for each learning procedure is defined as follows.

- Action choice of LP1 and LP3.
  - 1) If there is no subjective Nash equilibrium action in his/her game, decision maker  $i$  adopts random action, since his/her current perception gives no information about the Nash equilibrium.
  - 2) If there is only one subjective Nash equilibrium action in his/her game, decision maker  $i$  adopts the subjective Nash equilibrium action.
  - 3) Otherwise, decision maker  $i$  selects one of subjective Nash equilibrium actions randomly and adopts it, since he/she has identical payoff over all the subjective Nash equilibria.
- Action choice of LP2.
  - 1) If all the decision makers in  $A$  perceive  $S_B$  identically, decision maker  $i$  adopts action choice of LP1 and LP3.
  - 2) Otherwise, each decision maker adopts random action.

Action choice of LP2 assumes that as long as at least one decision maker has different perception about the nature's set of strategies, each decision maker has no strong belief in the current subjective Nash equilibria; accordingly, he/she applies random action.

### D. Fitness Evaluation

The fitness function maps his/her perceptions into an element in  $[0, 1]$ . The perceptions are evaluated by considering admissible information so far. We set costs in the function to measure the difference between the information and perceptions. The smaller the costs of a perception of the nature's game, the smaller the differences between the perception and the information.

- Fitness evaluation of LP1 and LP2.

Fitness evaluation for both the learning procedures employs a fitness function  $f_1$  which takes into account only preference-oriented information, i.e.,  $f_1$  evaluates perceptions by calculating two costs  $OC$  and  $CE_1$ .

The following is the procedure to calculate fitness score of a perception by  $f_1$ .

- 1)  $OC$  is opportunity cost defined by

$$OC = (|S_A| - i\text{'s current payoff}).$$

Accordingly,  $OC$  is a measure of the difference between  $|S_A|$ , the highest payoff possibly achieved by a decision maker, and his/her real payoff from the current game.

- 2)  $CE_1$  is cost of exchange representing the cost needed to change current perception about the nature's preference on the current result of the game into  $|S_{Bi}|$ ,  $i$ 's current perception of the highest payoff possibly achieved by nature by the definition of preference in Section II. Consequently,  $CE_1$  is based on the assumption that the nature has the highest preference on the current result of the game. That is, let  $(a_m, b_k)$  be the current result of the game between  $i$  and nature,  $P_{a_m}^{Bi}(a_m, b_k)$  be  $i$ 's perception of the nature's preference on the result, then  $CE_1$  is of the form

$$CE_1 = |S_{Bi}| - P_{a_m}^{Bi}(a_m, b_k).$$

- 3)  $f_1$  is of the form

$$f_1 = \frac{1}{1 + (w \times OC \times CE_1)}$$

where  $w$  is an index number defined in the next section, which represents the importance of  $CE_1$  in calculating the score of a perception by  $f_1$ .

- Fitness evaluation of LP3.

LP3 employs two fitness functions,  $f_1$  and  $f_2$ ; for each iteration, it adopts one of the following depending on current population of perceptions.

- 1) If for each  $i \in A$ ,  $|S_{Bi}|$  is identical to the number of different responses from the nature so far, then each decision maker adopts  $f_1$  defined as above. Accordingly,  $f_1$  in LP3 is used if all the decision makers in  $A$  has achieved an agreement on the perception of  $S_B$ . Since  $f_1$  only considers preference-oriented information, LP3 uses it to learn preference ordering of nature.
- 2) If not, then each decision maker employs  $f_2$ . Since  $f_2$  only considers strategy-oriented information, LP3 uses it to learn the nature's set of strategies.

Accordingly, at each iteration,  $i$  employing LP3 always evaluates the current population of perceptions before choosing one of the two fitness functions. By switching one from the other, LP3 considers both of preference- and strategy-oriented information.

LP3 consists of two phases of learning cyclically. The first phase is to learn the nature's set of strategies supported by  $f_2$ . The latter is to learn the nature's preference ordering supported by  $f_1$ . But from the phase 2, the learning may go back to the phase 1 again, if there is a new response applied by the nature.

$f_2$  calculates fitness score of  $i$ 's present perception by a cost  $CE_2$  measuring the difference between  $|S_{Bi}|$  and his/her forecast on  $|S_B|$  in the following way.



- 1)  $i$  calculates his/her forecast on  $|S_B|$  by observing the number of variations of the nature's responses so far as follows.

a) His/her subjective probability distribution function,  $h_i$ : set of all the possible number of elements of  $S_{Bi} \rightarrow [0, 1]$ , is defined:

Let  $E = \{1, 2, \dots, |S|\}$  be set of all the possible number of elements of  $S_{Bi}$ , and  $\eta$  be the number of different responses from the nature during the iterations so far. We will represent how many times the present  $\eta$  have been observed by  $i$  during the iterations so far by  $q_i(\eta)$ . Then let  $h_i: E \rightarrow [0, 1]$  be such that

$$h_i(x) = \begin{cases} 0 & \text{if } x < \eta \\ \frac{q_i(\eta)}{q_i(\eta) + (|S| - \eta)} & \text{if } x = \eta \\ \frac{1}{q_i(\eta) + (|S| - \eta)} & \text{otherwise} \end{cases}$$

$h_i$  is a current probability distribution function where  $\eta$  and  $q_i(\eta)$  are changeable in time horizon, since it depends on the response applied by nature in each iteration. As time goes on, if the variation of responses applied by the nature increases,  $\eta$  will get closer and closer to  $|S_B|$ . If the nature applies a new variation of response, then  $\eta$  increases, and by the definition of  $h_i$ , the probability of  $|S_B|$  to be smaller than the current  $\eta$  is zero. And, if the current  $\eta$  remains in the next iteration, then  $q_i(\eta)$  increases, so that it increases the probability of the current  $\eta$  to be  $|S_B|$ ; in contrast, for any  $y > \eta$ , it decreases the probability of  $y$  to be  $|S_B|$ .

b)  $i$ 's forecast on  $|S_B|$ , denoted by  $\psi_i$ , is a member in  $E$  selected randomly with respect to  $h_i(x)$  for each  $x \in E$  as follows.

Let  $Rand$  be a random number between 0 and 1, then  $i$  selects  $x \in E$  as  $\psi_i$  if  $\sum_{l=1}^x h_i(l) \geq Rand$  and there is no  $y \in E$ ,  $y < x$  such that  $\sum_{l=1}^y h_i(l) \geq Rand$ .

By this method,  $i$  tends to choose  $x \in E$  having higher probability to be his/her forecast on  $|S_B|$  than the smaller one.

- 2)  $CE_2$  is defined by

$$CE_2 = \text{the absolute value of } (|S_{Bi}| - \psi_i).$$

- 3) Finally,  $f_2$  is defined by

$$f_2 = \frac{1}{w \times CE_2 + 1}$$

where  $w$  is an index defined in the next section, that represents the importance of  $CE_2$  in calculating the fitness score of a perception by  $f_2$ .

#### E. Breaking Rule

Our learning procedures will stop when all the decision makers in  $A$  perceive the nature's behavior identically or the maximum number of iterations is reached. The identical (or final) perception here represents that the agreement of all the decision makers in  $A$  on the perception of the nature's game has been achieved; while the time to achieve the agreement is limited by the maximum number of iterations. If the agreement has been reached, the update process of the GA will be not able to produce new perceptions from the population of perceptions. According to the decision makers, the final perception represents the best perception that is possibly produced by the GA.

#### F. Selection

Selection in this paper is to choose a perception, called parent, from the population of  $n$  perceptions, when each decision maker begins to revise his/her perception. All three learning procedures use a selection method called "roulette wheel" referring to the ratio of each perception fitness score to the total fitness score of all the present perceptions, that is described as follows.

- Generate a random number between 0 and 1, called  $Rand$ .
- Let  $Fit(i)$  be fitness score of  $i$ 's perception, then calculate  $RatioFit(i) = (Fit(i) / \sum_{k=1}^n Fit(k))$ , for each decision maker  $i \in A = \{1, 2, \dots, n\}$ .
- Select perception of  $j \in A$  as a parent if  $\sum_{k=1}^j RatioFit(k) \geq Rand$  and there is no  $x \in A$ ,  $x < j$  such that  $\sum_{k=1}^x RatioFit(k) \geq Rand$ .

Accordingly, selection implies that all the decision makers in  $A$  share current perceptions and their fitness scores, in order to survive good perceptions of the nature's game.

After doing selection, each decision maker in the group begins to create children by reproduction or crossover. The group adopts a mechanism that decides whether a pair of decision makers adopt crossover or reproduction as follows. First, it generates a random number  $u \in [0, 1]$ . Then, if  $u < p_c$ , the pair adopt crossover; otherwise they adopt reproduction.

#### G. Reproduction

Reproduction is used by decision makers to produce new perceptions simply by copying the perception chosen by them in the selection.

#### H. Crossovers

Though we basically use a crossover method called "one point crossover," this paper modifies it to our purpose. The conventional crossover produces new chromosomes only by combining parents while crossover in this paper, in addition to combining parents, also revises chromosomes deliberately as soon as there is additional information about the nature's game. We define three crossovers called CO1, CO2, and CO3 for LP1, LP2, and LP3, respectively, where any gene in the chromosomes is identified with its position in the matrix described by Fig. 4. For example, the gene  $g_{mk}^r$  in  $C_r$  is identified as the gene located in the  $m$ th line and  $k$ th column [or located in the cell  $(m, k)$ ].

- Procedure of CO1.

Let the two parents selected by two decision makers  $i$  and  $j$  be  $C_r$  and  $C_s$  (chromosomes of decision makers  $r$  and  $s$  in  $A$ , respectively), and the fitness score of  $C_r$  be higher than  $C_s$ . And, letting the nature's current response be  $b_k$ , we define crossover as follows.

- 1) Select a cut point; i.e., if  $b_k \in S_{Br}$  holds, select the gene  $g_{|S_A|k}^r$  in  $C_r$  as the cut point. Otherwise, the cut point is selected randomly among the genes in the  $|S_A|$ th line.  $S_{Br}$  denotes perception of  $S_B$  decoded from  $C_r$ . This cut point is selected to maintain  $b_k$  in  $S'_{Br}$ , perception of  $S_B$  decoded from a chromosome  $C'_r$  that is resulted from the revision of  $C_r$  by combining it with  $C_s$ .
- 2) Combine the parents, that is  $C_r$  and  $C_s$  are combined by exchanging the genes located in the right hand of the cut point. We call chromosomes produced by the combination  $C'_r$  and  $C'_s$ . For example, let  $S = \{b_1, b_2, b_3\}$ ,  $S_A = \{a_1, a_2, a_3\}$ ,  $C_r$  and  $C_s$  be 121 000 212 and 211 122 000, respectively, and the cut point be  $g_{31}^r$ , then the combination is illustrated by Fig. 6.

The example produces  $C'_r$  and  $C'_s$  having a different perception about the nature's set of strategies and pref-

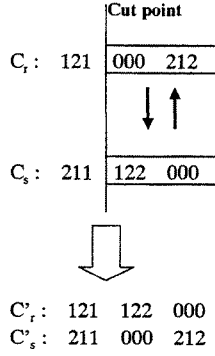


Fig. 6. Illustration of chromosome combination.

erence from the parents. As we can see from Fig. 6, the parents  $C_r$  and  $C_s$  imply, respectively  $S_{B_r} = \{b_1, b_3\}$  and  $S_{B_s} = \{b_1, b_2\}$ ; while the combination produces  $S'_{B_r} = \{b_1, b_2\}$  and  $S'_{B_s} = \{b_1, b_3\}$  implied by  $C'_r$  and  $C'_s$ , respectively. Accordingly, we know that if both the parents imply identical perceptions of  $S_B$ , then the combination will not produce new perceptions about  $S_B$ .

3) Revise the chromosomes produced by the combination.

We define how  $i$  performs these conscious revisions, while revisions by  $j$  are defined similarly. These revisions performed by  $i$  and  $j$  change  $C'_r$  and  $C'_s$  into  $C''_r$  and  $C''_s$ , respectively.

- a) Let the current interaction result between  $i$  and nature be  $(a_m, b_k)$ ,  $i$  changes the gene in the cell  $(m, k)$  so that it is the highest among the genes in the  $m$ th line. It means that  $i$  believes that nature has the highest preference on the current result.

- b) Restore the perception of nature's preference ordering implied by chromosomes produced by the revisions so far to the ordinal payoff function defined in Section II as follows.

Let for each  $a_m \in S_A$ ,  $c$ : {nonzero genes in the  $m$ th line}  $\rightarrow$

{1, 2, ..., number of nonzero genes in the  $m$ th line} be an ordinal function; such that for each pair of the nonzero genes in the cells  $(m, k)$  and  $(m, l)$  in the  $m$ th line,

- i) if the gene in  $(m, k) =$  the gene in  $(m, l)$  holds,  $c(\text{the gene in } (m, k)) > c(\text{the gene in } (m, l))$  or  $c(\text{the gene in } (m, k)) < c(\text{the gene in } (m, l))$ ,
- ii) if the gene in  $(m, k) >$  the gene in  $(m, l)$  holds,  $c(\text{the gene in } (m, k)) > c(\text{the gene in } (m, l))$ .

Accordingly,  $c$  induces a new perception of the nature's preference ordering implied by  $C'_r$  that is close to the old one, and is consistent with the ordinal payoff function defined in Section II. Then for each nonzero gene in the  $m$ th line,  $i$  replaces the gene with  $c(\text{the gene})$ .

• Procedure of CO2

CO2 consists of four steps: the first two steps of which are identical to the first two steps of CO1. Step 3) of CO2 is to adjust perception of  $S_B$  implied by chromosomes produced by the parents' combination to  $M_B$ , a set of all the different responses from the nature so far; while the last is to revise preference ordering implied by the chromosomes produced by the preceding step, which is defined identically to Step 3) of CO1. We assume

that all the decision makers in  $A$  share information about the different responses from the nature so far to set  $M_B$  in the current iteration. Step 3) of CO2 is defined as follows.

Let  $C'_r$  be a chromosome that  $i$  gets from the parents' combination, the following revisions are done by  $i$  in turn to adjust perception about  $S_B$  implied by  $C'_r$  to  $M_B$ ;

- a) For each response  $b_k \in M_B$ , if there is a zero gene in the  $k$ th column, the gene is changed into 1; otherwise there is no revision.
- b) For each response  $b_l \in S - M_B$ , if there is a nonzero gene in the  $l$ th column, the gene is changed into 0; otherwise there is no revision.

The chromosome  $i$  produces from this revision is called  $C''_r$ . This revision is also performed by  $j$  to produce  $C''_s$  from  $C'_s$  in the similar way.

• Procedure of CO3.

CO3 also consists of four steps: the first two steps of which are identical to the first two steps of CO1. Step 3) of CO3 consists of two cyclic stages of revisions related to perception of the nature's set of strategies. The last step of CO3 is to revise preference ordering implied by chromosomes produced by the preceding step, which is defined identically to Step 3) of CO1.

Step 3) of CO3 supports the two cyclical phases of learning of LP3, the first phase of which is to learn  $|S_B|$  supported by Step 3) of CO3 and  $f_2$ . The latter is to learn the nature's preference supported by  $f_1$ , after each decision maker revises his/her perception of  $S_B$  such that the perception is identical to  $M_B$  by using Step 3) of CO3.

Let  $C'_r$  be a chromosome that  $i$  gets from the parents combination, the following is the formal definition of Step 3) of CO3 that produces  $C''_r$ .

- a) If perceptions of  $|S_B|$  implied by all the current chromosomes are equal to  $|M_B|$ ,  $i$  adjusts  $S'_{B_r}$  to  $M_B$  by using Step 3) of CO2. That is, after the group of decision makers has agreed with each other that the true perception of  $|S_B|$  should be equal to  $|M_B|$ ,  $i$  changes perception of  $S_B$  implied by  $C'_r$  into  $M_B$ .
- b) If not,  $i$  adjusts the perception of  $|S_B|$  implied by  $C'_r$  to  $\psi_i$  (or  $i$ 's forecast on  $|S_B|$  defined in the fitness evaluation of LP3). The adjustment is defined as follows:
  - i) If  $|S'_{B_r}| < \psi_i$  holds,  $i$  adds some responses to  $S'_{B_r}$  as many as  $\psi_i - |S'_{B_r}|$  in the following way.  $i$  randomly selects  $\psi_i - |S'_{B_r}|$  elements from  $S - S'_{B_r}$ , then for each  $b_k$  of the selected elements,  $i$  replaces each zero gene in the  $k$ th column with one element of  $\{1, 2, \dots, \psi_i\}$  such that it is different from the other genes located in the same line as the gene.
  - ii) If  $|S'_{B_r}| > \psi_i$  holds,  $i$  eliminates some responses of  $S'_{B_r}$  as many as  $|S'_{B_r}| - \psi_i$  in the following way.  $i$  selects randomly  $|S'_{B_r}| - \psi_i$  elements from  $S'_{B_r}$ , then for each  $b_k$  of the selected elements,  $i$  replaces each gene in the  $k$ th column with zero.
  - iii) if  $|S'_{B_r}| = \psi_i$ ,  $i$  simply copies  $C'_r$  to  $C''_r$ .

Similarly,  $j$  using LP3 also follows Step 3) of CO3 to change  $C'_s$  into  $C''_s$ .

## V. DESIGN AND EXECUTION OF EXPERIMENTS

We test perform LP1, LP2, and LP3 in clarifying the nature's set of strategies and the outcomes most preferred by nature by simulating each of them in the hypergame situation, of which the general definition was given in Section II. We call the group of decision makers

successful in clarifying the nature's set of strategies and  $n$  outcomes most preferred by nature if at the last iteration, the population of chromosomes converge on a final chromosome that implies a correct perception about the nature's set of strategies and the outcomes.

#### A. Setting of the Experiments

We perform two experiments in this paper. The first experiment is to analyze the effects of uncertainty level of the nature's game on the performance of the learning procedures. That is, we provide several variations of  $|S|$  and  $|S_A|$  in the hypergame situation that changes the length of chromosomes, i.e.,  $|S_A| \times |S|$ .

The variation of  $|S|$  influences the uncertainty level; since the more elements of  $S$  exist, the more the alternatives of nature's set of strategies perceived by decision maker increase. Also, the increase of elements of  $S$  imposes the decision makers to consider more possible outcomes most preferred by the nature.

We provide seven alternatives of  $S$  of the hypergame situation with the fixed  $S_A = \{a_1, a_2, a_3\}$  and  $p_c = 0.6$  to analyze the effect of variations of  $|S|$  on the learning procedures; namely,  $\{b_1, b_2, b_3, b_4\}$ ,  $\{b_1, b_2, b_3, b_4, b_5\}$ , up to  $\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}\}$ , respectively.

Similarly, variations of  $|S_A|$  change the length of chromosome, so that they increase the uncertainty toward the outcomes most preferred by nature. For this purpose, we also set seven alternatives of  $S_A$  with the fixed  $S = \{b_1, b_2, b_3, b_4, b_5\}$  and  $p_c = 0.6$ ; namely,  $\{a_1, a_2, a_3\}$ ,  $\{a_1, a_2, a_3, a_4\}$ , up to  $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$ .

On the other hand, the second experiment is performed to analyze the effect of crossover rate (or  $p_c$ ) on the performances of the learning procedures. In this paper, crossover rate represents the rate of perception exchanges and the rate of conscious revisions as soon as there is additional information. That is, we set eight variations of crossover rate of the learning procedures with the fixed  $S_A = \{a_1, a_2, a_3\}$  and  $S = \{b_1, b_2, b_3, b_4, b_5\}$  to analyze the effect of crossover rates on the learning procedures; namely, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, respectively.

For all the experiments nature has fixed set of strategies  $S_B = \{b_1, b_3, b_4\}$ , and nature has  $|S_B|$  most preferred responses over  $S_A$ . As defined in Section II, each decision maker has no information about the nature's set of strategies and preference.

#### B. Setting of Control Parameters

Based on the results of presimulations, we set the other parameters for all the experiments as follows. Number of decision makers in the group is 30, number of maximum iteration is 1000, and  $w$  (index of both  $f_1$  and  $f_2$ ) is five.

#### C. Performances of Learning Procedures

We simulate each of the learning procedures in the hypergame situation for each alternative of the parameters  $|S|$ ,  $|S_A|$ , and crossover rate, as many as 30 runs. Then, the average performance of each learning procedure in learning about the nature's behavior will be compared.

Fig. 7 illustrates how each of final chromosomes C-lp1, C-lp2, and C-lp3 evolves in a simulation of, respectively, LP1, LP2, and LP3 in the hypergame situation with  $|S_A| = 3$ ,  $|S| = 6$ , and the outcomes most preferred by nature are  $(a_1, b_1)$ ,  $(a_2, b_3)$ , and  $(a_3, b_4)$ .

In the simulation of LP1, the population of perceptions converges on C-lp1 at iteration 70th; while in the simulations of LP2 and LP3, the population converges on C-lp2 and C-lp3 at iterations 44<sup>th</sup> and 93<sup>rd</sup>, respectively. By decoding C-lp1, we know that in the simulation, LP1 fails in clarifying the nature's set of strategies, but it is able to clarify two of the three outcomes most preferred by nature. Perception of the nature's set of strategies implied by C-lp1 is  $\{b_1, b_2, b_4, b_5, b_6\}$ ,

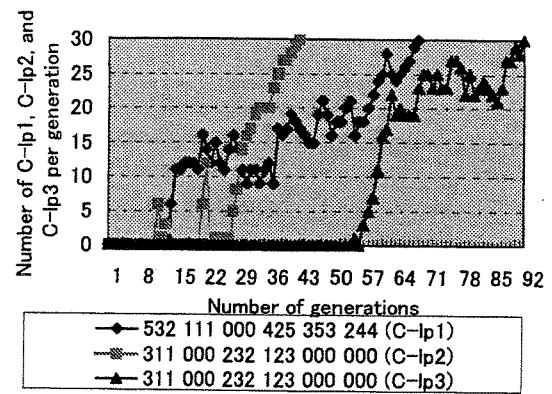


Fig. 7. Increase of final perceptions by each learning procedure.

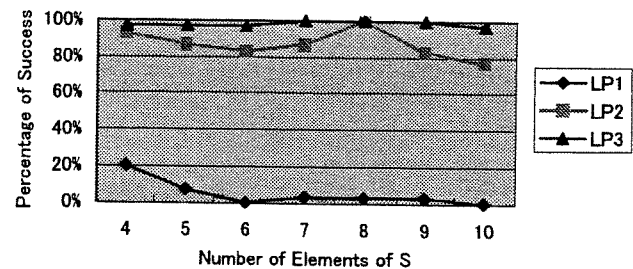


Fig. 8. Effect of  $|S|$  on learning the nature's set of strategies.

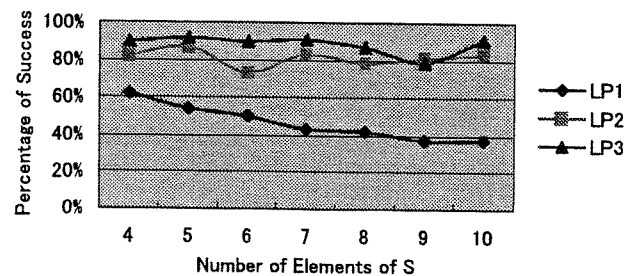


Fig. 9. Effect of  $|S|$  on learning the outcomes most preferred by the nature.

while perception of two outcomes most preferred by nature implied by C-lp1 are  $(a_1, b_1)$  and  $(a_3, b_4)$ . On the other hand, by decoding C-lp2 and C-lp3, we know that in the simulation, LP2 and LP3 succeed in clarifying either the nature's set of strategies or all the three outcomes most preferred by the nature.

#### 1) Effects of Uncertainty Levels on the Learning Procedures:

- Effects of variations of  $|S|$  on the learning procedures

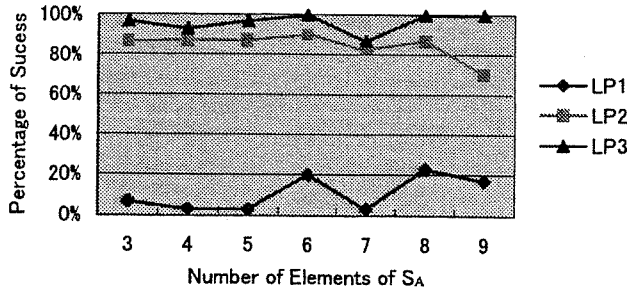
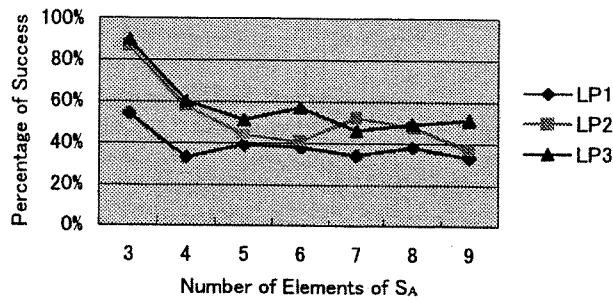
The effects are described by Figs. 8 and 9.

As seen in Figs. 8 and 9, LP1 has poor performance in clarifying the nature's set of strategies, while its performance in clarifying the outcomes most preferred by nature decreases as  $|S|$  increases. That is, when  $|S| = 4$  LP1 can clarify in average 62% of all the outcomes most preferred by nature, but it decreases until 37% when  $|S| = 10$ . In contrast, LP2 and LP3 mostly succeed in clarifying the nature's set of strategies, has better performance in learning the outcomes most preferred by the nature, while the performances are not influenced by the increase of  $|S|$ .

- Effects of Variations of  $|S_A|$  on the Learning Procedures

Figs. 10 and 11 describe the effects.

The effect of variations of  $|S_A|$  on the learning procedures in clarifying the nature's set of strategies is similar with the effect of  $|S|$  on them as seen in Figs. 8 and 10. However, the performance of LP2 tends to decrease as  $|S_A| > 6$ , and the increase of  $|S_A|$

Fig. 10. Effect of  $|S_A|$  on learning the nature's set of strategies.Fig. 11. Effect of  $|S_A|$  on learning the outcomes most preferred by the nature.

does not weaken the performance of LP1 in clarifying the nature's set of strategies.

Fig. 11 describes that performances of all the learning procedures in clarifying the outcomes most preferred by nature decrease as  $|S_A|$  increases.

The variations of either  $|S|$  or  $|S_A|$  influence the length of the chromosome, and they represent variations of uncertainty level of the nature's behavior. That is, by the definition of preference of the both parties in Section II, for each  $a_m \in S_A$ , there is  $b_k \in S$  such that  $(a_m, b_k)$  is one of the outcomes most preferred by nature, that is the probability of an outcome to be one of the outcomes most preferred by nature is  $(1/|S|)$ . Accordingly, the probability that a decision maker correctly perceives all the outcomes most preferred by nature is  $((1/|S|))^{|S_A|}$ . Also, the variations of  $|S|$  influence uncertainty toward the nature's set of strategies, since  $S_{Bi}$  is a subset of  $S$ .

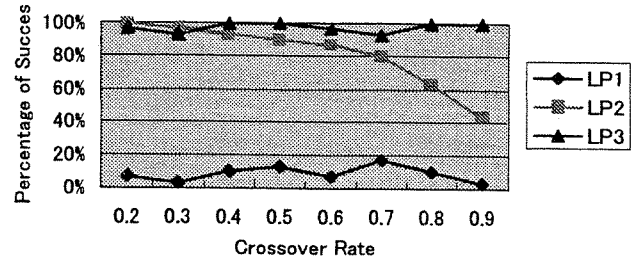
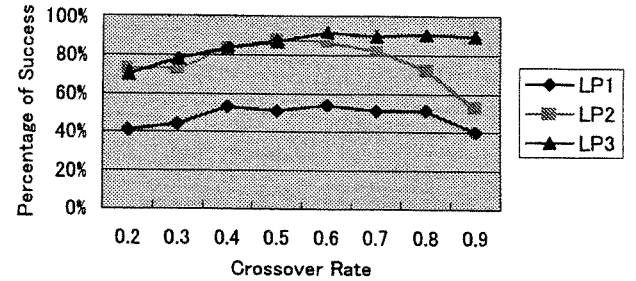
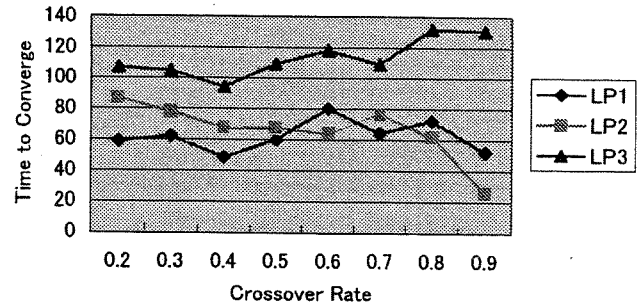
2) *Effects of Crossover Rate on the Learning Procedures:* They are described by Figs. 12 and 13.

As seen in Figs. 12 and 13, similar to the previous experiments, LP1 has poor performance in clarifying the nature's set of strategies and the outcomes most preferred by nature, while LP2 and LP3 have better performance in clarifying them than LP1.

However, performance of LP2 in learning the nature's set of strategies is influenced by the variations of crossover rate. That is, its performance decreases as  $p_c$  increases.

On the other hand, until  $p_c = 0.6$  the increase of crossover rate improves the performance of all the learning procedures in learning the outcomes most preferred by nature. However, as  $p_c > 0.6$ , the increase of crossover rate decreases the performance of LP1 and LP2; while the performance of LP3 tends to be constant.

The increase of  $p_c$  gives a different effect on the speed of convergence of the learning procedures as seen in Fig. 14. The speed of convergence of LP1 tends to be constant even though  $p_c$  increases; while LP3 needs more time to converge as  $p_c$  increases. In contrast to LP3, the time to converge of LP2 decreases as  $p_c$  increases.

Fig. 12. Effect of  $p_c$  on learning the nature's set of strategies.Fig. 13. Effect of  $p_c$  on learning the outcomes most preferred by the nature.Fig. 14. Effect of  $p_c$  on the speed of convergence.

#### D. Discussions

These experiments show that the performance of the learning procedures in clarifying all the outcomes most preferred by nature is closely related to the performance in clarifying the nature's set of strategies.

The performance differences especially between LP1 and the rest show that the different definitions of the learning procedures in at least one activity of action choice, fitness evaluation, and crossover have a significant effect on the learning procedures' performances. We summarize the differences of the learning procedures by Table I.

LP1 emphasizes only preference-oriented information contained in Action Choice, Fitness Evaluation, and Crossover. It tends to be unable to improve perceptions about the nature's set of strategies; as well as to clarify all the outcomes contained in the nature's game and also the nature's preference ordering over the outcomes. Accordingly, as the uncertainty level of the nature's set of strategies increases, the performance of LP1 in clarifying the nature's set of strategies decreases; and in all the experiments it has constantly poor performance in clarifying the nature's set of strategies and preference.

LP2 and LP3 are proposed to eliminate the LP1's weaknesses. Although LP2 uses LP1's fitness evaluation, it employs action choice and crossover accommodating strategy-oriented information so far.

Experiment shows that conscious revisions in CO2 is mostly able to clarify the nature's set of strategies, since it changes perception about the nature's set of strategies into the common  $M_B$ , set of various responses applied by nature so far. Accordingly, the performance of LP2 in learning the nature's set of strategies is not influenced by  $|S|$ , as seen

TABLE I  
DIFFERENCES OF THE LEARNING PROCEDURES

Proc.	Some activities of the procedures		
	Action choice	Fitness evaluation	Crossover
LP1	depending on $N_i$ , use sub. Nash equilibrium act. or random action	use $f_1$	1. select cut point 2. combine parents 3. revise preference ordering
LP2	depending on $\alpha^t$ , use sub. Nash equilibrium act. or random action	use $f_1$	1. select cut point 2. combine parents 3. adjust percept. of $S_B$ to $M_B$ 4. revise preference ordering
LP3	depending on $N_i$ , use sub. Nash equilibrium act. or random action	depending on $\alpha^t$ , use $f_1$ or $f_2$	1. select cut point 2. combine parents 3. depending on $\alpha^t$ , adjust perception of $S_B$ to $M_B$ , or perception of $ S_B $ to $\psi_i$ 4. revise preference ordering

in Fig. 8. LP2 has also better performances in clarifying the nature's set of strategies and the outcomes most preferred by nature than LP1 in all the experiments.

However, since CO2 changes each  $S_{Bi}$  into  $M_B$ , the success of LP2 in clarifying the nature's set of strategies depends on its ability to force nature to apply all its responses. The increase of  $|S_A|$  especially when  $|S_A| > |S_B|$ , decreases the ability, because it increases possibility of the nature to apply the same response to various actions of a decision maker. Accordingly, the conscious change likely forces the population of chromosomes to converge quickly on a final chromosome that implies misperception about the nature's set of strategies, as seen in Fig. 10. Moreover, the increase of crossover rate of CO2 reduces variation of population of perceptions in shorter time as in Fig. 14. So, it produces more chances for LP2 to produce an incorrect perception of the nature's set of strategies as seen in Fig. 12, and to lower its performance in clarifying the outcomes most preferred by nature as  $p_c > 0.6$  as described by Fig. 13.

On the other hand, LP3 employs LP1's action choice, but it employs fitness evaluation and crossover accommodating strategy-oriented information. LP3 consists of two phases of learning, of which the first phase is to learn  $S_B$  by employing  $f_2$  and CO3 revising  $|S_{Bi}|$  so that it is identical to  $\psi_i$  (or  $i$ 's subjective forecast on  $S_B$ ) until each decision maker's perception about  $|S_B|$  is equal to  $|M_B|$ , and then revising  $S_{Bi}$  by CO3 so that it is identical to  $M_B$ . The latter is to clarify the nature's preference by employing  $f_1$ .

The experiments show that two phases of learning considering strategy- and preference-oriented information cyclically is mostly able to clarify the nature's set of strategies; and the performance tends to be constant in all the experiments.

By the two phases of learning of LP3, rather than changing all the perceptions of the nature's set of strategies into  $M_B$  from the beginning, LP3 delays the process until the decision makers has achieved an agreement on perceptions of  $|S_B|$  by using a conscious change based

on individual perceptions, that is Step 3) of CO3. Accordingly, CO3 maintains variation of the population of perceptions, and gives enough time for LP3 to learn the nature's set of strategies. As seen in Figs. 14 and 13, the increase of  $p_c$  does not lower the time to converge of LP3 and its performance in clarifying the outcomes most preferred by nature.

However, the performance of all the learning procedures in clarifying the outcomes most preferred by nature decreases as  $|S_A|$  increases; because the increase of  $|S_A|$  augments the number of outcomes most preferred by nature, and by definition of preference in Section II there are some lines in the subjective game impossibly contain a subjective Nash equilibrium when  $S_A > S_B$ . Then, by the subjective Nash equilibrium action, the result of the game with nature can not exist in the lines. Accordingly, the conscious revision in all the learning procedures (or Step 3 in CO1 and Step 4 in CO2 and CO3) is difficult to clarify the outcomes most preferred by nature in the lines.

## VI. CONCLUSIONS

This paper proposes three types of learning procedures called LP1, LP2, and LP3, for a group of decision makers to improve perceptions about the nature's set of strategies and preferences using a GA. Then, we performed two experiments to analyze the effect of uncertainty levels and crossover rates on the performances of the learning procedures in clarifying the nature's set of strategies and preferences.

Experimental results showed that

- LP1 adopting only preference-oriented information has poor performance in clarifying either the nature's set of strategies or all the outcomes most preferred by the nature. And its performance decreases as the uncertainty increases.
- LP2 was proposed to remove LP1's weakness by accommodating a conscious change that produces identical perceptions about the nature's set of strategies based on common strategy-oriented

information, and using action choice and crossover based on both of preference- and strategy-related information. LP2 has better performance in clarifying the nature's set of strategies and preferences than LP1. However, its performances in clarifying the nature's set of strategies and preferences decrease quickly as crossover rate  $> 0.6$ .

- LP3, using two cyclic phases of learning, was designed to overcome the weakness of LP2, that is to delay the conscious change that produces identical perceptions until the group has finished learning about the number of elements of the nature's set of strategies by a conscious change based on individual beliefs of decision makers considering strategy-oriented information. It also accommodates both preference- and strategy-related information in fitness evaluation and crossover. Our experiment showed that the two phases of learning can maintain variations of population of perceptions even though the crossover rate  $> 0.6$ . In general, LP3 has better performance in learning the nature's set of strategies and preferences than LP2.

Finally, we believe that our research provides a new approach to understand the nature's behavior better and to harmonize our decision with the nature. Some applications to real decision situations are now under way.

#### APPENDIX LIST OF SYMBOLS AND NOTATION

$1, \dots, i, \dots, n$	symbols for the decision makers;
$A = \{1, 2, \dots, n\}$	the group of decision makers;
$B$	symbol for nature;
$S_A = \{a_1, a_2, \dots, a_{ S_A }\}$	the common set of strategies for $A$ ;
$S_B = \{b_1, b_2, \dots, b_{ S_B }\}$	set of strategies of nature;
$S$	universal set of strategies;
$S_{Bi}$	$i$ 's perception of $S_B$ ;
$P_{b_j}^A: S_A \times \{b_j\} \rightarrow \{1, 2, \dots,  S_A \}, \forall b_j \in S$	$A$ 's common preference ordering on $S_A \times \{b_j\}$ ;
$p_{jk} = P_{b_j}^A(a_k, b_j)$	$A$ 's common preference over $(a_k, b_j)$ ;
$P_{a_k}^{Bi}: \{a_k\} \times S_{Bi} \rightarrow \{1, 2, \dots,  S_{Bi} \}, \forall a_k \in S_A$	$i$ 's perception of the nature preference over $\{a_k\} \times S_{Bi}$ ;
$q_{kj}^i = P_{a_k}^{Bi}(a_k, b_j)$	$i$ 's perception of the nature's preference over $(a_k, b_j)$ ;
LP1, LP2, and LP3	learning procedures;
$\alpha^t = (\alpha_1^t, \alpha_2^t, \dots, \alpha_n^t)$	a population of perceptions at the $t$ th iteration;
$g_{mk}^i = g_i(a_m, b_k)$	a gene for a situation $(a_m, b_k)$ ;
$C_i$	the chromosome of $i$ ;
$C_i^t, C_i''$	chromosomes after revisions in crossover;
$N_i$	Set of subjective Nash equilibria in $i$ 's subjective game;
$f_1, f_2$	fitness functions;
$OC$	opportunity cost in fitness function;

$CE_1, CE_2$

$E$

$\eta$

$M_B$

$q_i(\eta)$

$h_i: E \rightarrow [0, 1]$

$p_c$

$\psi_i$

cost of exchanges in  $f_1$  and  $f_2$ , respectively;  
set of possible number of elements of  $S_{Bi}$ ;  
number of different responses from nature so far;  
set of different responses from nature so far;  
frequency of present  $\eta$  observed by  $i$  so far;  
probability distribution function on  $E$ ;  
crossover rate;  
 $i$ 's forecast on number of elements of  $S_B$ .

#### REFERENCES

- [1] D. E. Goldberg, *Genetic Algorithm: In Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
- [2] G. J. Klir, *Facets of Systems Sciences*. New York: Plenum, 1991.
- [3] H. A. Simon, "Information technologies and organizations," *The Accounting Review*, vol. 65, no. 3, pp. 656–667, 1990.
- [4] J. Rosenhead, *Rational Analysis for a Problematic World*. New York: Wiley, 1989.
- [5] K. A. Dowsland, "Genetic algorithms—A tool for OR?," *J. Oper. Res. Soc.*, vol. 47, pp. 550–561, 1996.
- [6] K. Kijima, "An intelligent poly-agent learning model and its application," *Inform. Syst. Eng.*, vol. 2, pp. 47–61, 1996.
- [7] —, "Poly-agent system theory" (in Japanese), in *Proc. Jpn. Soc. Manag. Inform. Conf.*, June 1997, pp. 57–60.
- [8] L. Davis, *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold, 1991.
- [9] M. O. Odetayo, "Knowledge acquisition and adaptation: A genetic approach," *Expert Syst.*, vol. 12, no. 1, pp. 3–13, 1995.
- [10] M. Wang, "Strategic learning in games with incomplete information," *IEEE Int. Conf. Syst., Man, Cybern.*, vol. 3, pp. 2047–2052, 1996.
- [11] M. Wang, K. W. Hipel, and N. M. Fraser, "Solution concepts in hypergames," *Appl. Math. Comput.*, vol. 34, no. 3, pp. 147–171, (1989).
- [12] P. B. Checkland, *Systems Thinking, Systems Practice*. New York: Wiley, 1981.
- [13] P. G. Bennett, "Hypergames: Developing a model of conflict," *Futures*, vol. 12, no. 6, pp. 489–507, 1980.

## **SIMULATION OF ADAPTATION PROCESS IN HYPERGAME SITUATION BY GENETIC ALGORITHM\***

UTOMO SARJONO PUTRO<sup>a</sup>, KYOICHI KIJIMA<sup>b,†</sup>  
and SHINGO TAKAHASHI<sup>c</sup>

<sup>a</sup>*Department of Value and Decision Science, Tokyo Institute of Technology,  
Tokyo 152, Japan, e-mail: Utomo@valdes.titech.ac.jp;*

<sup>b</sup>*Department of Value and Decision Science, Tokyo Institute  
of Technology, Tokyo 152, Japan;*

<sup>c</sup>*Department of Management and Systems Engineering, Chiba Institute  
of Technology, Chiba 275, Japan, e-mail: Tshingo@cc.it-chiba.ac.jp*

*(Received 23 July 1999)*

The purpose of this paper is to analyze adaptation process of two interacting groups of agents in hypergame situation by simulation, where each group adopts an internal learning behavior. That is, we attempt to clarify which learning behavior facilitates the adaptation process to converge on equilibria of the traditional game situation (TGS), and facilitates the groups to learn the equilibria correctly. First, we define the hypergame situation, in which each agent is assumed to have only internal model of the situation and exchanges information only with other agents in the same group. Then, we develop adaptation process model, in which genetic algorithm has role to improve each group's perceptions, and a simulation of the process. Finally, by examining the simulation results, we point out that learning behavior accommodating subjective Nash equilibria is critical to the performance of the adaptation process, in the situations with one or more TGS Nash equilibria.

**Keywords:** Hypergame; genetic algorithm; adaptation process; bounded rationality; learning.

---

\*An earlier version of this paper was presented at IEEE International Conference on Systems, Man and Cybernetics held in Tokyo, October 12–15, 1999. This work was in part supported by Grant-in-Aid for Scientific Research (A) 11694025 from the Japanese Government.

<sup>†</sup>Corresponding author. e-mail: Kijima@valdes.titech.ac.jp

## 1. INTRODUCTION

The purpose of this paper is to analyze adaptation process of two interacting groups of agents in hypergame situation by simulation, where each group adopts an internal learning behavior. That is, we attempt to clarify which learning behavior facilitates the adaptation process to converge on equilibria of the traditional game situation (TGS), and facilitates the groups to learn the equilibria correctly.

Much of the traditional economists' works were devoted to the question of how the equilibrium behavior changes when parameters in the model change, but only very little to the problem how the transition from one equilibrium to another actually takes place [6]. One of the reason is they assume that economic system is always in equilibrium. To state an economic system will *a priori* in equilibrium, they have to assumpt that all agents in the system are willing to and able to maximize their expected utility, and have rational expectation. Rational expectation in game theoretic model means that all agents in the system have identical and exactly correct beliefs about how every agent will behave.

Different from the traditional economists, our research deals with adaptation process of bounded rational agents. We relax the assumption of rational expectation, since it does not always reflect reality; *i.e.*, we assume that the agents may have different beliefs about the situation and how every agent will behave. Accordingly, there is no guarantee that the adaptation process always converges on a equilibrium. One of motivations of this paper is to clarify adaptation process that facilitates interacting group of agents to converge towards equilibria.

Recently, economics also deals with adaptation process of bounded rationality, relying on plausible assumptions about agents, it provides several mathematical learning models to the adaptation process of bounded rationality; namely least square learning [1,2], stochastic approximation [12] and Bayesian learning [10]. However, these models mostly assume a representative agent who represents a whole population of agents. The representative agent updates his belief and reacts to his expectation according to a learning rule [6].

On the other hand, artificial intelligence (AI) especially computational intelligence (CI) provides learning techniques such as neural



networks, genetic algorithms, and classifier systems based on genetic algorithms or cellular automata. They try to deal with learning models based on heuristic arguments. Supported by the learning techniques and the capability of modern computer, a learning model, that allows us to describe adaptation process of individual agents each of which behaves according to his/her internal learning behavior and model of the situation, can be developed. That is, by learning techniques of CI we can analyze more complex model of the adaptation process than by mathematical learning techniques.

Application of the learning models of economics and AI, in game theory is in evolutionary game theory, which deals with learning in iterated games.

Existing researchs that deal with adaptation process of bounded rationality in evolutionary game theory are based on the traditional game theory. In the TGS, each agent has common model of situation; *i.e.*, each agent's preference and set of strategies is known perfectly by any other agents.

Our research deals with adaptation process in a situation where bounded rationality is more clearly expressed than in TGS. We analyze adaptation process of two interacting groups where each agent in a group is assumed to have common set of strategies and preference and may have misperception about the other group's common set of strategies and preference. The interactive situation is called hypergame situation [14]. In this situation, each agent is supposed to have internal model of the situation, based on which he/she makes decisions. We will formulate the internal model as a subjective game in this paper.

In this paper, adaptation process of a group of agents is described as follows. Each agent in the group plays a game with a different agent in the other group. Given a current perception of the other group's common set of strategies and preference, he/she chooses an action based on his/her group's learning behavior. Strategies applied by the other group so far are the only information about the group that can be observed by each agent. Besides the information, each agent may consider internal information, for instances, his/her current payoff and positions of subjective Nash equilibria to judge perception. Based on the judgment and information exchange with some of agents in the same group, each agent revises his/her perception. In the next game, he/she uses the new perception. By iterating the game and revision

again and again, each agent may have better and better perception about the situation.

In the ideal case, both groups in a long run have common model of the situation; *i.e.*, the situation turns into TGS. However, the main purpose of our adaptation process is to facilitate each group to learn only the equilibria of TGS, though both of the groups may still have misperceptions about some aspects of the situation.

We develop an adaptation process model, in which genetic algorithm (GA) is used to improve population of perceptions. GA maintains a population of solutions based on past experiences to search good solutions. It is suitable for our purpose that deals with modification of a population of perceptions according to the past experiences to search perceptions that correctly perceive TGS Nash equilibria.

In organizations, random adaptive processes may occur occasionally, but we can also make a conscious decision that certain things contribute to survival and other things don't and, thus, deliberately affect modification of the system [5]. Accordingly, we adopt two kind of changes in our adaptation process, namely random and conscious changes.

In this paper, we modify GA so that besides random changes, it also accommodates conscious changes.

We simulate the adaptation process in several representations of the hypergame situation. In a scenario of the simulation, both the group adopt the same learning behaviors; and in another scenario they adopt different learning behaviors.

Learning behaviors, each group adopts one of them, define what information is necessary to judge perception and how actions are chosen. Technically, they represent internal mechanisms, by which each group deliberately guide the adaptation process. We propose two learning behaviors, namely LB1 and LB2. LB1 assumes that each agent chooses the most preferred action to the preceding action of his/her rival. On the other hand, LB2 assumes that each agent's action is based on his/her subjective Nash equilibria.

After formulating hypergame situation we consider in Section 2, we will explain idea of adaptation process model in Section 3. Section 4 develops a simulation of the adaptation process. Section 5 describes the results of the simulation and then discuss the results and findings. Section 6 gives the conclusions.

## 2. HYPERGAME SITUATION

First of all we define hypergame situation handled by this paper as follows:

- It consists of two groups of agents where each agent in a group has common preference and set of strategies. The groups play a game with each other, in which agent  $i$  in a group is always paired with his/her rival  $j$  (or an agent in the other group).
- Each group is uncertain with the other group's common preference, and only knows a universal set containing the other group's common set of strategies. Accordingly, each agent can not calculate equilibria of TGS; he/she only has perceptions of the equilibria.
- Each agent is assumed to have an internal model of interaction between the groups, so that the interaction between a pair of agents is determined by a pair of internal models.
- The internal model of an agent may be shared with some agents in the same group but it is not shared with any agent in the different group.

As an illustration of the hypergame situation described above, we provide a case of two firms,  $A$  and  $B$ , that compete on the same market. The firms may produce several types of products for the market, and each firm has a distributor in each region. Product types produced by the firms are based on their distributors' demands.

Let  $i \in A$  and  $j \in B$  be a pair of interacting distributors in a region.  $S_A$  (similarly  $S_B$ ) denotes common set of strategies of  $A$  (similarly  $B$ ).  $i$  (similarly  $j$ ) has incomplete information about  $S_B$  (similarly  $S_A$ ); that is  $i$  (similarly  $j$ ) only knows a universal set  $U_{S_B}$  (similarly  $U_{S_A}$ ) containing  $S_B$  (similarly  $S_A$ ). The elements of  $U_{S_A}$  (similarly  $U_{S_B}$ ) represents product types possibly sold in the market by  $i$  (similarly  $j$ ).  $S_{Bi}$  and  $S_{Aj}$  denote  $i$ 's perception of  $S_B$  and  $j$ 's perception of  $S_A$  respectively.  $\pi_A(\alpha, \beta)$  (similarly  $\pi_B(\alpha, \beta)$ ) represents common preference of  $i$  (similarly  $j$ ) over any outcome  $(\alpha, \beta) \in S_A \times S_{Bi}$  (similarly  $S_{Aj} \times S_B$ ). Perception of  $i$  (similarly  $j$ ) about  $\pi_B(\alpha, \beta)$  (similarly  $\pi_A(\alpha, \beta)$ ) is denoted by  $\pi_{Bi}(\alpha, \beta)$  (similarly  $\pi_{Aj}(\alpha, \beta)$ ).

Then, the hypergame situation that describes interaction between  $i$  and  $j$  in a region is illustrated by Figure 1. The figure shows that each agent is only able to build perceptions of the situation.

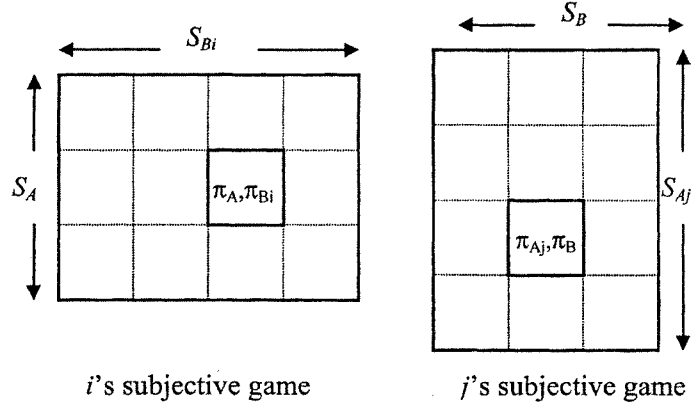


FIGURE 1 Illustration of the hypergame situation.

### 3. ADAPTATION PROCESS MODEL

This paper proposes adaptation process of both the groups, each of which adopts a learning behavior in the hypergame situation. We simulate the adaptation process to know the effects of learning behaviors, LB1 and LB2, on facilitating the adaptation process to converge toward TGS Nash equilibria, and on facilitating the groups to learn the equilibria.

LB1 assumes that any agent in each group chooses his/her most preferred action to the preceding action of his/her rival. He/she considers information about  $B$ 's responses to each action of  $A$  so far to evaluate current perceptions. On the other hand, LB2 assumes that each agent's action is based on his/her subjective Nash equilibria. He/she considers positions of subjective Nash equilibria in his/her subjective game to evaluate current perceptions.

The adaptation process model of each group is described as follows:

1. *Stage 1 Initialization:* Generate randomly an initial population of  $n$  perceptions, each of which is associated with an agent in the group. Each perception is encoded by a bit string (or a chromosome).
2. *Stage 2 Action choice:* Given his/her own current perception, each agent in the group plays a game with an agent in the other group. The way the agent selects an action is defined by a learning behavior adopted by his/her group.

### 3. Stage 3 Update:

(a) Revise the current population of perceptions by GA as following procedure:

- Fitness evaluation: Calculate fitness score of each perception in the current population with a fitness function. What information that is accommodated by the function depends on which learning behavior adopted by the group.
- Breeding process: Repeat the procedure below until produce  $n$  children (or  $n$  new perceptions produced from the current perceptions).
  - Selection: Select 2 perceptions from the current population as parents, in such a way that the fitter perceptions have the higher chances to be chosen as parents, and a perception may be chosen more than one.
  - Reproduction or crossover:
    - \* Generate random number  $u$  from Uniform (0,1) distribution
    - \* If  $u < p_c$ , do crossover: the mating of the two parents to produce two new perceptions. Where  $p_c$  is crossover probability defined in Table I. The probability is determined by pre-simulation results.
    - \* If not, do reproduction: two children are produced by simply copying the two parents.

(b) Replace the current population of perceptions with the population of  $n$  children. The flowchart of the adaptation process model of each group is described in Figure 2.

GA “mimics” the natural selection process to evolve good individuals in a population of individuals [9]. The natural selection is a

TABLE I Definition of parameters

Number of agents in each group (or $n$ )	30
Number of maximum iterations	2000
Probability of crossover (or $p_c$ )	60%
Probability of reproduction	40%
$w_1$	15
$w_2$ and $w_3$	5

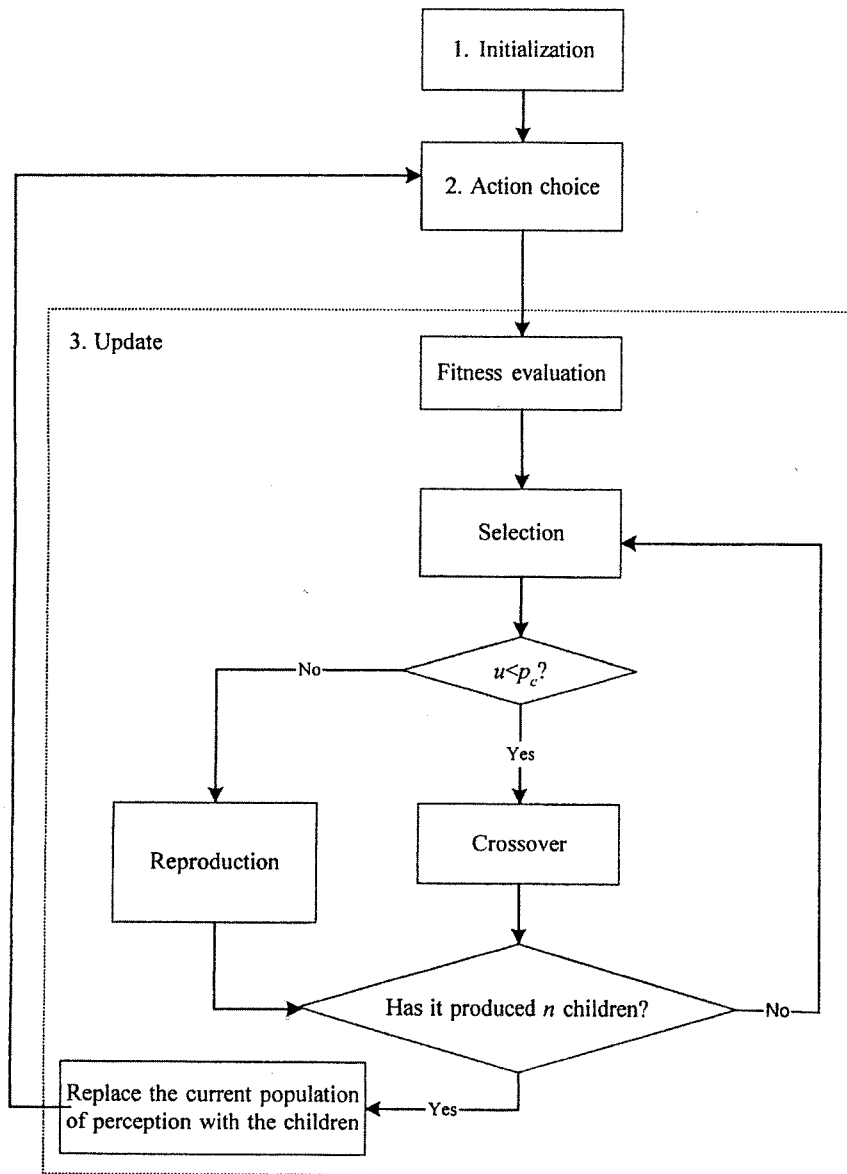


FIGURE 2 Flowchart of the adaptation process model.

mechanism which survives the fittest individual in some population where each individual competes each other for, say, food or mating. The fitter individuals tend to win the competition, and they have more chances to be chosen as parents in the breeding process than the unsuccessful individuals. During the breeding process, there may happen reproduction, crossover and mutation. Some children have

new characteristics better than their parents, produced by crossover and mutation; while the other children have characteristics simply copied from parents. By repeating the competition again and again, the population will be dominated by better and better individuals.

The adaptation process model above does not employ mutation, because it provides only small local revision, that is not enough for our purpose. We also modify crossover in such a way that it is able to adjust perceptions to additional information about the other group's strategies applied so far. Accordingly, crossover in this paper also represents conscious changes.

#### 4. A SIMULATION OF THE ADAPTATION PROCESS

The simulation model of the adaptation process of the groups is described as follows:

1. *Step 1 Initialization*: Execute Stage 1 of the learning model for both the groups simultaneously.
2. *Step 2 Playing games*: Execute Stage 2 of the learning model for both the groups simultaneously.
3. *Step 3 Termination*: If the perception population of each group has converged, or the maximum number of iteration has been reached, then STOP. Otherwise, go to Step 4. The maximum number is defined in Table I.
4. *Step 4 Update*: Execute Stage 3 of the learning model for the group whose perception population has not converged yet.
5. *Step 5 Games repetition*: Go to Step 2.

A perception produced by the adaptation process of a group is the perception on which a perception population of the group has converged.

##### 4.1. Assumptions of the Hypergame Situation

We define preference of each agent  $i$  in group  $A$  over any outcome as follows:  $i$  has common preference over any outcome in  $S_A \times S_{Bi}$ , namely for each  $\beta \in S_{Bi}$ ,  $i$  employs a common ordinal utility function  $\pi_A: S_A \times \{\beta\} \rightarrow \{1, 2, \dots, |S_A|\}$ , where  $\pi_A$  is bijective. Also,  $i$ 's perception of  $B$ 's preference is represented by an ordinal utility function

$\pi_{Bi}$  for each  $\alpha \in S_A$  and each  $\beta \in S_{Bi}$  defined by  $\pi_{Bi}: \{\alpha\} \times S_{Bi} \rightarrow \{1, 2, \dots, |S_{Bi}|\}$ , where  $\pi_{Bi}$  is bijective.

In the similar way, we define  $B$ 's preference over any outcome.

We define a Nash equilibrium of TGS as follows.

$(\alpha^*, \beta^*) \in S_A \times S_B$  is called a Nash equilibrium of TGS if  $(\forall \alpha \in S_A) (\pi_A(\alpha^*, \beta^*) \geq \pi_A(\alpha, \beta^*))$  and  $(\forall \beta \in S_B) (\pi_B(\alpha^*, \beta^*) \geq \pi_B(\alpha^*, \beta))$ .

We denote a set of all the Nash equilibria of TGS by  $N$ .

For the hypergame situation, we define subjective Nash equilibria in internal model of each agent  $i$  in the group  $A$  as follows:  $(\alpha^*, \beta^*) \in S_A \times S_{Bi}$  is called a subjective Nash equilibrium if  $(\forall \alpha \in S_A) (\pi_A(\alpha^*, \beta^*) \geq \pi_A(\alpha, \beta^*))$  and  $(\forall \beta \in S_{Bi}) (\pi_{Bi}(\alpha^*, \beta^*) \geq \pi_{Bi}(\alpha^*, \beta))$ .

We denote a set of all the subjective Nash equilibria in  $i$ 's game by  $N_i$ .

Accordingly,  $i$  correctly perceives a TGS Nash equilibrium  $(\alpha^*, \beta^*) \in N$  if  $(\alpha^*, \beta^*) \in N \cap N_i$ .

We call the adaptation process of both the groups successful in converging towards equilibria of TGS and the groups can learn the equilibria if at the last iteration, each adaptation process of any pair of interacting agents,  $i \in A$  and  $j \in B$ , converges on an outcome belonging to  $N$  and both the agents correctly perceive it as a subjective Nash equilibrium.

#### 4.1.1. Representations of the Situation

Each representation of the hypergame situation is of the form  $(A, U_{S_B}, S_A, \pi_A, B, U_{S_A}, S_B, \pi_B)$ . We consider 15 different representations which are grouped into five packets, namely  $P_1, P_2, P_3, P_4$  and  $P_5$ . Every packet consists of three different representations which have one, two, and three TGS equilibria respectively.  $|U_{S_A}|$  and  $|U_{S_B}|$  in the representations vary from 5 to 10. And the representations define differently  $S_A, \pi_A, S_B$  or  $\pi_B$ , but have the same  $|S_A| = |S_B| = 3$ . Number of agents in each group is defined in Table I.

The next subsections define components of the learning model.

#### 4.2. Encoding

Encoding is to represent an agent's perception about the other group's common preference and set of strategies by a bit string or



a chromosome. We define encoding for each agent  $i$  in group  $A$ , while the encoding definition for group  $B$  can be defined similarly.

Let  $S_A = \{a, c, d\}$  where  $a$ ,  $c$  and  $d$  are located in respectively the first, second and third lines of  $i$ 's subjective payoff matrix, and  $U_{S_B} = \{p, q, r, s\}$ . With each  $(\alpha, \beta) \in S_A \times U_{S_B}$ , we associate a function  $g_i(i = 1, 2, \dots, n)$  defined by  $g_i: \{(\alpha, \beta) | (\alpha, \beta) \in S_A \times U_{S_B}\} \rightarrow \{0, 1, 2, \dots, |S_{Bi}|\}$ , which is such that for each  $\alpha \in S_A$ ,

$$g_i(\alpha, \beta) = \begin{cases} \pi_{Bi}(\alpha, \beta) & \text{if } \beta \in S_{Bi} \\ 0 & \text{otherwise} \end{cases}$$

Then  $i$ 's chromosome  $C_i$  is a string consisting of the genes as long as  $|S_A| \times |U_{S_B}|$  and is of the form described by Figure 3.

Zero genes in  $C_i$  represent all outcomes that do not exist in  $i$ 's subjective game. On the other hand, non zero genes represent  $i$ 's perception of the opponent's common preference over all outcomes in  $i$ 's subjective game.

For example, let an agent  $r \in A$  has  $S_{Br} = \{p, r\}$ , then  $C_r$  is described by Figure 4.

### 4.3. Action Choices

An action choice of an agent represents how he/she chooses actions given his/her perception of the situation. This paper defines two kinds of action choice, namely the best response to the rival's preceding action and subjective Nash equilibria action, for LB1 and LB2 respectively.

$$\underbrace{g_i(a, p)g_i(c, p)g_i(d, p)}_{S_A \times \{p\}} \underbrace{g_i(a, q)g_i(c, q)g_i(d, q)}_{S_A \times \{q\}} \underbrace{g_i(a, r)g_i(c, r)g_i(d, r)}_{S_A \times \{r\}} \underbrace{g_i(a, s)g_i(c, s)g_i(d, s)}_{S_A \times \{s\}}$$

FIGURE 3 Agent  $i$ 's chromosome.

$$\underbrace{\pi_{Br}(a, p)\pi_{Br}(c, p)\pi_{Br}(d, p)}_{S_A \times \{p\}} \underbrace{\pi_{Br}(a, r)\pi_{Br}(c, r)\pi_{Br}(d, r)}_{S_A \times \{r\}} \underbrace{\pi_{Br}(a, s)\pi_{Br}(c, s)\pi_{Br}(d, s)}_{S_A \times \{s\}}$$

FIGURE 4 Agent  $r$ 's chromosome.

Let  $A$  adopts LB1, then for each  $i \in A$ ,  $i$  adopts the best response to the preceding action of his/her rival as follows.

$\alpha^* \in S_A$  is called the best response to  $j$ 's preceding action if for all  $\alpha \in S_A$ ,  $\pi_A(\alpha^*, \beta^o) \geq \pi_A(\alpha, \beta^o)$ , where  $\beta^o$  is  $j$ 's action at the preceding iteration.

On the other hand, if  $A$  adopts LB2, then for each  $i \in A$ ,  $i$  adopts the subjective Nash equilibria action as follows.

$\alpha^* \in S_A$  is called a subjective Nash equilibria action if there is  $\beta^* \in S_{B_i}$  such that  $(\alpha^*, \beta^*) \in N_i$ .

#### 4.4. Fitness Evaluation

The fitness evaluation implies how agent  $i$  calculates the fitness of his/her perception by using a fitness function. The fitness function maps his/her perception into an element in  $[0, 1]$ .

The definition of fitness function depends on which learning behavior adopted by  $i$ 's group. However, in general an agent sets the fitness score of his/her own perception by evaluating its relevance to some admissible information so far, and/or payoff achieved by him/her [16]. If current perception is inconsistent with the information, there are costs to modify the perception. The costs lower the fitness of the perception. The more the inconsistency is the smaller the perception fitness becomes.

##### 4.4.1. Fitness Function of LB1

Let  $A$  adopts LB1, then each agent  $i$  in the group evaluates his/her perception fitness score by a fitness function called  $f_1$ .  $f_1$  evaluates  $i$ 's perception by taking into account current payoff, and  $B$ 's responses to each action of  $A$  so far.

By LB1,  $A$  assumes that each agent  $j \in B$  also chooses the best response to the preceding action of his/her rival  $i$ . Agents in  $A$  exchange information each other about their rivals' responses to their preceding actions so far. Based on the information,  $i$  calculates  $B$ 's dominant response to each action of  $A$ , where the response will be interpreted by  $i$  as the most preferred response of  $B$  to the action of  $A$ .

$f_1$  maps  $i$ 's perception to a fitness score by calculating two costs  $OC_1$  and  $CE$ ;

1.  $OC_1$  is opportunity cost defined by

- if  $j$ 's current action is in  $S_{Bi}$ ,

$$OC_1 = (\text{the highest payoff} - i\text{'s current payoff}).$$

- if not,  $OC_1 = \text{the highest payoff}$ .

Accordingly, if  $i$  is perceiving correctly his/her rival  $j$ 's current action, then  $OC_1$  is the difference between the highest payoff possibly achieved by  $i$  and his/her current payoff. Otherwise,  $OC_1$  is the highest payoff.

2.  $CE$  is cost of exchange representing the cost needed to change current perception with respect to  $B$ 's responses to each action of  $A$  so far, defined as following way:

- (a) Let  $\alpha \in S_A$  be an action applied by members of  $I \subset A$  at iteration  $t - 1$ , and  $J \subset B$  is set of rivals of  $I$ 's members. And  $n_{\beta}^t$  be number of agents in  $J$  that give response  $\beta$  to  $\alpha$  at iteration  $t$ . Then the frequency of  $\beta$  applied as response to  $\alpha$  until the current iteration  $T$  (or so far) is calculated by  $F_{\alpha\beta} = \sum_{t=1}^T n_{\beta}^t$ . At the current iteration, for each  $\beta \in S_{Bi}$ ,  $i$  calculates  $F_{\alpha\beta}$  and finds  $\beta^* \in S_{Bi}$  where for any  $\beta \in S_{Bi}$ ,  $F_{\alpha\beta^*} \geq F_{\alpha\beta}$ .  $\beta^*$  is called  $B$ 's dominant response to  $\alpha$  so far. LBI assumes that  $\beta^*$  is the most preferred response of  $B$  to  $\alpha$ .

(b)  $CE$  is defined as follows.

- (i) Let  $\forall \beta \in S_{Bi}$ ,  $\pi_{Bi}(\alpha, \beta') \geq \pi_{Bi}(\alpha, \beta)$ , if  $\beta' \neq \beta^*$  then set  $CE_{\alpha} = 1$ ; otherwise  $CE_{\alpha} = 0$ . That is, if  $i$  is perceiving that  $\beta^*$  is the most preferred response of  $B$  to  $\alpha$  then there is no need to modify his/her current perception (or  $CE_{\alpha} = 0$ ); otherwise he/she needs to modify it (or  $CE_{\alpha} = 1$ ).
- (ii) We calculate  $CE = \sum_{\alpha \in S_A} CE_{\alpha}$ . That is,  $CE$  is total cost related with  $i$ 's perception of the most preferred response of  $B$  to all different actions of  $A$ .

3.  $f_1$  is of the form

$$f_1 = \frac{1}{1 + (w_1 \times OC_1 + w_2 \times CE)}.$$

Where  $w_1$  and  $w_2$  are index numbers defined in Table I, which are employed to adjust effect of the costs on  $f_1$ . The numbers are determined by pre-simulation results.

#### 4.4.2. Fitness Function of LB2

Let  $A$  adopts LB2, then each agent  $i$  in the group evaluates his/her perception fitness score by a fitness function called  $f_2$ .  $f_2$  assumes that each agent chooses actions based on his/her own subjective Nash equilibria.  $f_2$  maps  $i$ 's perceptions to a fitness score by calculating  $OC_2$  that accommodating information about subjective Nash equilibria and current payoff:

1. If  $\beta^\circ \in S_{Bi}$  holds,  $i$  calculates an opportunity cost  $oc$  = (the highest payoff  $-i$ 's current payoff), where  $\beta^\circ$  is  $j$ 's current action. Otherwise,  $oc$  = the highest payoff.
  - If  $oc = 0$ ,  $i$  believes that the current outcome  $(\alpha^\circ, \beta^\circ)$  of interaction between  $i$  and  $j$  is a TGS Nash equilibrium, by the assumptions of the hypergame situation and subjective Nash equilibria action of LB2.  $i$  identifies whether or not the outcome is in  $N_i$ . If yes,  $i$  believes that he/she is perceiving correctly a TGS Nash equilibrium (or  $OC_2 = 0$ ). Otherwise,  $i$  needs to modify his/her current perception, so that the current outcome is perceived by  $i$  as a subjective Nash equilibrium; *i.e.*,  $OC_2$  = perception of the highest payoff of  $B - \pi_{Bi}(\alpha^\circ, \beta^\circ)$ .
  - If not,  $i$  needs to apply another subjective Nash equilibria action; *i.e.*,  $OC_2 = oc$ .
2.  $f_2$  is of the form

$$f_2 = \frac{1}{1 + (w_3 \times OC_2)}.$$

Where  $w_3$  is index number defined in Table I, which is employed to adjust effect of the costs on  $f_2$ . The number is determined by pre-simulation results.

#### 4.5. Selection

Selection in this paper is to represent how each group chooses perceptions, called parents, from the population of perceptions. The

learning model uses a selection method called “roulette wheel” referring to the ratio of each perception fitness score to the total fitness score of a present population of  $n$  perceptions. We define how each agent  $i \in A = \{1, 2, \dots, n\}$  adopts “roulette wheel” selection as follows:

- Generate a random number  $Rand$  from Uniform  $(0, 1)$  distribution.
- Let  $Fit(r)$  be fitness score of  $r$ 's perception, then calculate  $RatioFit(r) = (Fit(r) / \sum_{k=1}^n Fit(k))$ , for each agent  $r \in A$ .
- $i$  selects perception of  $q \in A$  as a parent if there is no  $x \in A$ ,  $x < q$  such that  $\sum_{r=1}^x RatioFit(r) \geq Rand$ .

Similarly, we define how each agent in  $B$  adopts the selection method.

#### 4.6. Crossover

Crossover in this study represents how every two agents in a group combine their perceptions in order to produce two new perceptions. Though we basically use a crossover method called “one point crossover”, this paper modifies it to our purpose. The conventional crossover produces new chromosomes only by combining parents, while crossover in this paper not only combines parents, but also revises chromosomes with respect to the additional information about the other group's set of strategies. That is, we adopt both of random and conscious modification by crossover. We define crossover for  $A$ ; while crossover for  $B$  can be defined similarly.

Let the two parents selected by any two agents in  $A$ ,  $h$  and  $i$ , be  $C_r$  and  $C_s$  respectively, and fitness score of  $C_r$  be higher than  $C_s$ . And let  $S_{Br} = \{q, s\}$ ,  $U_{S_B} = \{p, q, r, s\}$  and  $S_A = \{a, c, d\}$  where  $a, c, d$  are respectively located in the first, second, and third lines of subjective payoff matrix of each agent in  $A$ . We denote the dominant action of  $B$  at the current iteration by  $\beta^o \in U_{S_B}$ . The chromosome of any agent in  $A$  is of the form described by Figure 3.

Then we define crossover as follows:

1. Select a cut point.

The cut point is determined so that if the fitter parent's perception of  $B$ 's set of strategies contains the dominant action of  $B$  at the current iteration, then crossover will maintain the action in the fitter parent's new perception of  $B$ 's set of strategies.

That is, if  $\beta^\circ \in S_{Br}$  holds, the cut point  $\zeta$  is a gene in both of the parents associated with outcome  $(d, \beta^\circ)$ , namely  $g_r(d, \beta^\circ)$  in  $C_r$  and  $g_s(d, \beta^\circ)$  in  $C_s$ . Otherwise,  $i$  selects randomly an element  $\beta$  in  $U_{S_B}$ , and the cut point  $\zeta$  is a gene in both of the parents associated with outcome  $(d, \beta)$ .

2. Combine the parents.

$h$  and  $i$  combine the parents  $C_r$  and  $C_s$  by exchanging each other the genes located in the right hand of  $\zeta$ . We call chromosomes produced by the combination  $C'_r$  and  $C'_s$ , which imply new perceptions of  $B$ 's preferences (or  $g'_r$  and  $g'_s$ ) and set of strategies (or  $S'_{Br}$  and  $S'_{Bs}$ ).

Let  $S_{Bs} = \{p, r\}$  and  $\beta^\circ = q$  then the combination is illustrated by Figure 5. The illustration of the combination produces  $C'_r$  and  $C'_s$  that imply perceptions of  $B$ 's set of strategies different from the parents  $C_r$  and  $C_s$ . As we can see from Figure 5, the parents  $C_r$  and  $C_s$  imply respectively  $S_{Br} = \{q, s\}$  and  $S_{Bs} = \{p, r\}$ , while the combination produces  $S'_{Br} = \{q, r\}$  and  $S'_{Bs} = \{p, s\}$  as being implied by  $C'_r$  and  $C'_s$  respectively. Accordingly, we know that if both the parents imply identical perception of  $S_B$ , then the combination will not produce new perception about  $S_B$ .

3. Revise the chromosomes produced by the combination.

This revision is to adjust the chromosomes to additional information about the other group's set of strategies as well as to adjust perception of  $B$ 's preference implied by them. We define this revision for  $C'_r$  by  $h$ ; while revision for  $C'_s$  by  $i$  can be defined similarly.

(a)  $h$  revises  $C'_r$  with respect to the additional information.

(i)  $h$  eliminates all the elements in  $S'_{Br}$  which are not contained in  $M_B$ , namely set of different actions applied by  $B$  so far, by  $(\forall \beta \in S'_{Br})(\neg(\beta \in M_B) \rightarrow (\forall \alpha \in S_A)(g'_r(\alpha, \beta) = 0))$ . We call the chromosome  $h$  produces from this revision  $C''_r$ .

(ii)  $h$  adds all the elements in  $M_B - S''_{Br}$  to  $S''_{Br}$  by  $(\forall \beta \in M_B)(\neg(\beta \in S''_{Br}) \rightarrow (\forall \alpha \in S_A)(g''_r(\alpha, \beta) = 1))$ .

The chromosome  $h$  produces from this revision is called  $C'''_r$ .

For example, let  $S'_{Br} = \{q, r\}$ ,  $M_B = \{r, s\}$ , and the genes associated with  $\{d\} \times U_{S_B}$  in  $C'_r$ ,  $g'_r(d, p)g'_r(d, q)g'_r(d, r)g'_r(d, s)$ ,

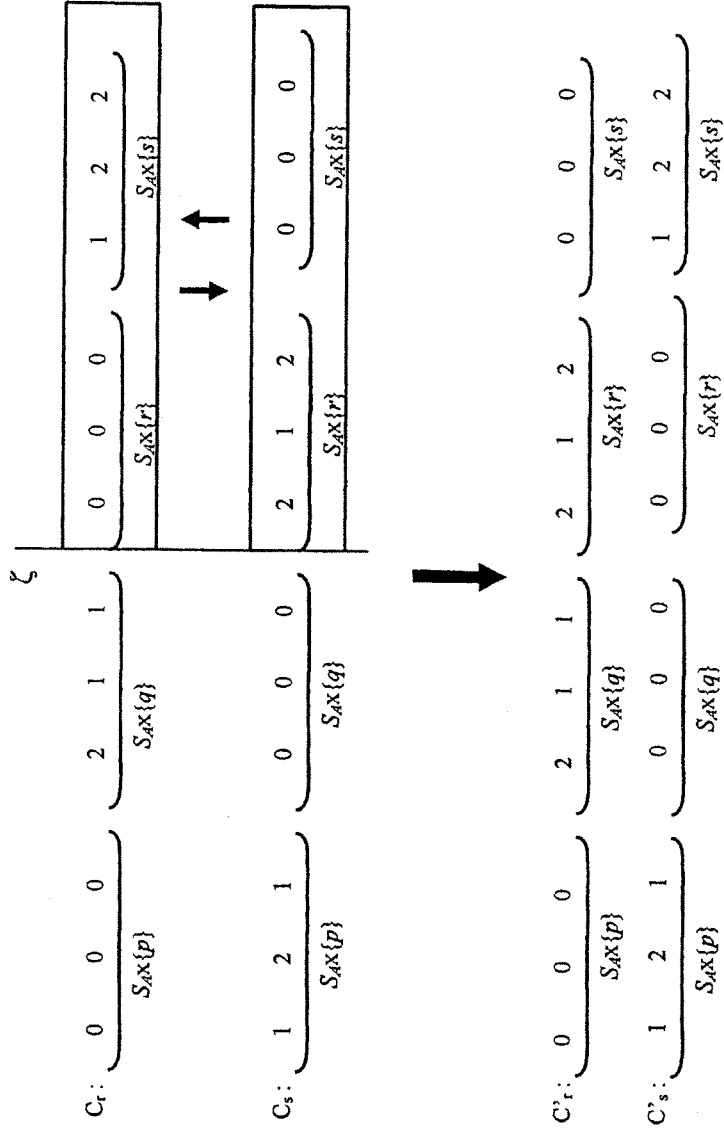


FIGURE 5 Illustration of chromosome combination.

be 0120 respectively. Then after this revision, the genes  $g_r'''(d,p)g_r'''(d,q)g_r'''(d,r)g_r'''(d,s)$  in  $C_r'''$  are 0021 respectively.

(b)  $h$  adjusts  $g_r'''$  to an ordinal utility function as follows.

Let for each  $\alpha \in S_A$ ,  $c : \{g_r'''(\alpha, \beta) | (\alpha, \beta) \in \{\alpha\} \times S_{Br}'''\} \rightarrow \{1, 2, \dots, |S_{Br}''|\}$  be an ordinal utility function where for any  $(\alpha, \beta), (\alpha, \gamma) \in \{\alpha\} \times S_{Br}'''$ ,

- if  $g_r'''(\alpha, \beta) = g_r'''(\alpha, \gamma)$  holds,  $c(g_r'''(\alpha, \beta)) > c(g_r'''(\alpha, \gamma))$  or  $c(g_r'''(\alpha, \beta)) < c(g_r'''(\alpha, \gamma))$ ,
- if  $g_r'''(\alpha, \beta) > g_r'''(\alpha, \gamma)$  holds,  $c(g_r'''(\alpha, \beta)) > c(g_r'''(\alpha, \gamma))$ .

$c$  induces a new perception of  $B$ 's preference that is close to the old one, and is consistent with the assumptions of the hypergame situation in Section 4. Then for each  $(\alpha, \beta) \in S_A \times S_{Br}'''$ ,  $h$  replaces  $g_r'''(\alpha, \beta)$  with  $c(g_r'''(\alpha, \beta))$ . We set the chromosome  $h$  produces in this way as  $C_r''''$ .

Let  $S_{Br}''' = \{r, s\}$ , and the values of genes associated with  $\{d\} \times U_{S_B}$  in  $C_r'''$ ,  $g_r'''(d,p)g_r'''(d,q)g_r'''(d,r)g_r'''(d,s)$ , be 0031 respectively. Then after this revision, the genes  $g_r''''(d,p)g_r''''(d,q)g_r''''(d,r)g_r''''(d,s)$  in  $C_r''''$  are 0021 respectively.

#### 4.7. Adaptation Process Scenarios

In each representation of the hypergame situation we simulate three scenarios of the adaptation process, namely SCN1-1, SCN1-2 and SCN2-2. In SCN1-1, both of  $A$  and  $B$  adopt LB1, in SCN1-2, one of them adopts LB1 and the other one adopts LB2, and in SCN2-2, both of them adopt LB2.

#### 4.8. Setting of Control Parameters

We set parameters for the simulation of the adaptation process as Table I, which are based on the results of pre-simulation.

### 5. SIMULATION RESULTS

In our experiments, we simulate each scenario of the adaptation process in any representation. In this section we describe and analyze the results of the simulation.



Figure 6 illustrates how scenarios SCN1-1 SCN1-2 and SCN2-2 of the adaptation process in  $R_{2NE}^4$  (or a representation in  $P_4$  that has two Nash equilibria of TGS) of the hypergame situation increase number of successful pairs.

The figure shows how number of successful pairs, as percentage of all the pairs, increases by time. As defined before, each group consists of 30 agents, each of which plays games with a different agent in the other group. A pair of interacting agents is called a successful pair if adaptation process of the pair converges on a TGS Nash equilibrium and both the agents correctly perceive it as a subjective Nash equilibrium.

We know from Figure 6 that in SCN1-1 of the adaptation process, perception populations of both the groups converge at iteration 22 (or the last iteration of the simulation of SCN1-1). While in SCN1-2 and SCN2-2 the populations converge at iteration 24 and 11 respectively.

In SCN1-1 of the adaptation process of both the groups, number of the successful pairs at the last iteration are only 57% of total pairs. Accordingly, in the illustration SCN1-1 fails in converging on TGS Nash equilibria and both the groups can not learn the equilibria.

In contrast, in SCN1-1 and SCN2-2 of the adaptation process, at the last iteration all the pairs are successful in converging on TGS Nash equilibria and can learn the equilibria. Accordingly, in the illustration SCN1-2 and SCN2-2 are successful in converging towards TGS Nash equilibria and both of the groups can learn the equilibria.

The simulation of each scenario of the adaptation process in a representation of the hypergame situation will be executed as many

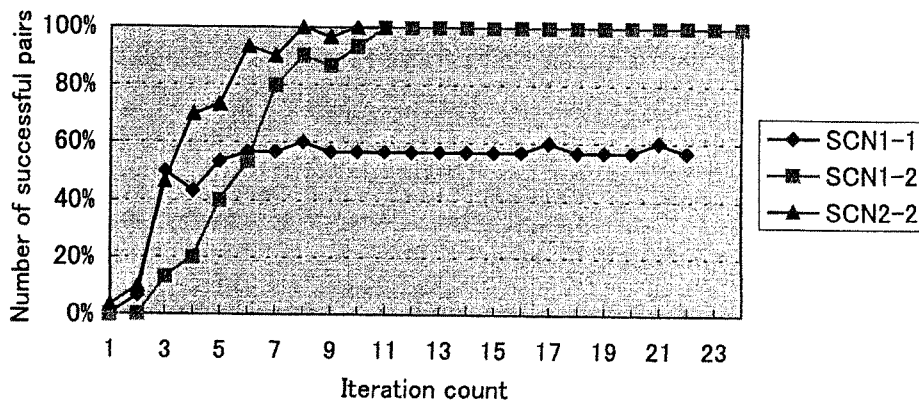


FIGURE 6 Illustration of each scenario of the adaptation process.

as 30 times. We want to know the performances of the scenarios in each representation. The performances represent number of the simulation executions that are successful in arriving at Nash equilibria of TGS and in learning the equilibria, stated in percentage of the 30 executions. Table II describes the performances. As we can see from Table II, except  $R_{1NE}^1$  (or a representation in  $P_1$  that has one Nash equilibrium of TGS) in all representations, SCN1-2 performance is better than SCN1-1. Also, except  $R_{1NE}^2$  in all packets SCN2-2 performance is better than performance of SCN1-1 and SCN1-2. And, generally in the representations with more than one Nash equilibrium (two or three equilibria), SCN2-2 is better than SCN1-2 and SCN1-1; while SCN1-2 is better than SCN1-1.

The performance of the scenarios related with the number of TGS Nash equilibria is resumed in Figure 7. As we can see in Figure 7, SCN2-2 and SCN2-1 have higher probability to converge on an Nash equilibrium of TGS and it can learn well it, though in situations with more than one equilibria than SCN1-1. In contrast, SCN1-1 may converge on an equilibrium only in the situations with one TGS Nash equilibrium, but most of agents fail in perceiving it as a subjective Nash equilibrium. While in the situations with more than one

TABLE II Performance of scenarios in each representation

<i>Packets</i>		<i>Scenarios</i>		
		SCN1-1	SCN1-2	SCN2-2
$P_1$	$R_{1NE}^1$	13%	10%	40%
	$R_{2NE}^1$	0%	80%	93%
	$R_{3NE}^1$	0%	57%	77%
$P_2$	$R_{1NE}^2$	57%	67%	43%
	$R_{2NE}^2$	0%	60%	77%
	$R_{3NE}^2$	0%	53%	90%
$P_3$	$R_{1NE}^3$	50%	77%	77%
	$R_{2NE}^3$	0%	60%	87%
	$R_{3NE}^3$	0%	40%	70%
$P_4$	$R_{1NE}^4$	40%	57%	67%
	$R_{2NE}^4$	0%	90%	93%
	$R_{3NE}^4$	0%	53%	60%
$P_5$	$R_{1NE}^5$	50%	63%	90%
	$R_{2NE}^5$	0%	37%	87%
	$R_{3NE}^5$	0%	50%	67%

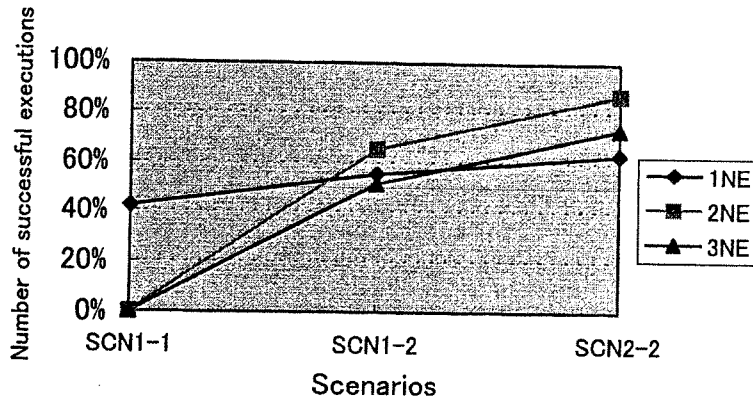


FIGURE 7 Performance of scenarios related with the number of TGS Nash equilibria.

TGS Nash equilibria, not all of games played by  $i$  and  $j$  converge on an equilibrium.

The results imply that action choice based on subjective Nash equilibria (or action choice of LB2) is critical to facilitate the adaptation process to converge towards an equilibrium. That is, it facilitates each game played by  $i$  and  $j$  to converge on the same equilibrium of TGS. GA has an important role on it, because GA will eventually make population of perceptions converge on a good perception. Accordingly, as time goes the number of agents in a group who have the perception increase. By subjective Nash equilibria choice, number of agents in the group who apply the same action increases. Similarly, if the other group also adopts LB2, agents in the group eventually apply the same action. On the other hand, if the other group adopts LB1, then number of agents in the group who apply the same action also increases by the best response to the rivals' preceding actions.

The results also imply that information about current payoff and positions of subjective Nash equilibria (in  $f_2$ ), leads GA to be able to survive perceptions that correctly perceive TGS Nash equilibria.

Different from LB2, action choice in LB1 does not rely on current perceptions, rather it relies on empirical data about the preceding rival's action. Accordingly, if the opponent also adopts LB1, then though current perception population has converged on a perception, agents in a group are still apply different actions. The simulation shows that when there are more than one equilibria of TGS, the

iteration of games played by  $i$  and  $j$  are difficult to converge on an equilibrium.

## 6. CONCLUSIONS

The present paper developed an adaptation process model of two interacting groups of agents in hypergame situation, where each group adopts a learning behavior. The learning behavior defines what information is necessary to judge perception (or internal model) of the situation, and how actions are chosen given the perception. We proposed two learning behaviors, LB1 and LB2. LB1 assumes that each agent chooses the best response to his/her rival's preceding action. On the other hand, LB2 assumes that each agent's action is based on his/her subjective Nash equilibria.

Then, to analyze the adaptation process we developed a simulation of the adaptation process, and repeated its executions in each of 15 representations of the hypergame situation as many as 30 times.

Simulation results showed that action choice based on subjective Nash equilibria (or action choice of LB2) is critical to facilitate the adaptation process to converge towards an equilibrium, and facilitate each group to learn the equilibria. Also, information about current payoff and positions of subjective Nash equilibria (in  $f_2$  of LB2) supports GA well to survive perceptions which correctly perceive TGS Nash equilibria.

In the most of the representations, if both of the groups adopt LB2, the performance of their adaptation process is better than if only one of the groups adopt LB2, and the performance of adaptation process when only one of the groups adopt LB2 is better than when both of the groups adopt LB1. Accordingly, without knowing which learning behavior adopted by the other group, it is better for a group to adopt LB2.

We believe that our research provides a new approach to deal with adaptation process in hypergame situation.

## References

- [1] A. Marcet, T. J. Sargent (1989). Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models, *Journal of Economic Theory*, **48**, 337–368.
- [2] C. A. Sims (1980). Macroeconomics and Reality, *Econometrica*, **48**(1), 1–48.

- [3] D. E. Goldberg (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, USA, Addison-Wesley Publishing Co.
- [4] G. J. Klir (1991). *Facets of Systems Sciences*, (International Federation for Systems Research International Series on Systems Science and Engineering, Vol. 7), New York, Plenum Press
- [5] H. A. Simon (1990). Information Technologies and Organizations, *The Accounting Review*, **65**(3), 656–667.
- [6] H. Dawid (1996). *Adaptive Learning by Genetic Algorithms: Analytical Results and Applications to Economic Models*, Berlin/Heidelberg, Springer-Verlag.
- [7] J. Rosenhead (1989). *Rational Analysis for a Problematic World*, New York, John Wiley & Sons Ltd.
- [8] K. Kijima (1996). An Intelligent Poly-Agent Learning Model and Its Application, *Information and System Engineering*, **2**(1), 47–61.
- [9] L. Davis (1991). *Handbook of Genetic Algorithms*, New York, Van Nostrand Reinhold.
- [10] L. E. Blume, D. Easley (1982). Learning To Be Rational, *Journal of Economic Theory*, **26**, 340–351.
- [11] M. O. Odetayo (1995). Knowledge Acquisition and Adaptation: A Genetic Approach, *Expert Systems*, **12**(1), 3–13.
- [12] M. Woodford, D. Easley (1990). Learning To Believe in Sunspots, *Econometrica*, **58**(2), 277–307.
- [13] P. B. Checkland (1981). *Systems Thinking, Systems Practice*, New York, John Wiley & Sons Ltd.
- [14] P. G. Bennett (1980). Hypergames: Developing a Model of Conflict, *Futures*, **12**(6), 489–507.
- [15] S. Takahashi, B. Nakano, M. Arase (1995). Application of Genetic Algorithm to Analysis of Adaptation Processes of Individual Perceptions in Hypergame Situation (in Japanese), *Journal of the Japan Society for Management Information*, **4**(1), 43–46.
- [16] U. S. Putro, K. Kijima, S. Takahashi, Adaptive Learning of Hypergame Situations by Using Genetic Algorithm, *IEEE Transactions on Systems, Man and Cybernetics* **30**(5), 562–572.

# エージェントベースモデリングのための数理的システム理論の課題

高橋 真吾

キーワード：エージェント,システム理論,遺伝的アルゴリズム,学習プロセス

## 1. はじめに

近年 agent-based modeling あるいは agent simulation と呼ばれる領域が発展している。これらはエージェントという概念に基づいて組織や経済システムにおける複雑な性質をシミュレーションにより解明しようとする[2,12,14]。そこではマルチエージェントシステムにおける分散人工知能等のエージェント技術も有効に利用される[9]。一方で複雑システムとしての創発的挙動を理解するための理論的基盤が十分になされているとはいえない。

システム理論は元来そのような複雑システムにおける創発的性質の理解に主眼が置かれてきたといっても過言ではない[16]。またそのためのシステムモデルがこれまでも開発されてきた。agent-based modeling における概念と方法は新たなシステムモデル開発の方向を与えてくれる可能性がある。また、agent-based modeling に対してもシミュレーション結果を構造的に説明する理論的基盤が提供されることはその発展にとり不可欠である。

本論文では、まず現在の数理的システム理論の位置をシステムの性質の階層的関係から整理し、次に agent-based modeling の特徴を述べた後、agent-based modeling における学習プロセスに対するシステムモデルの必要性和そのための数理的システム理論が直面している課題を考察する。

## 2. システム的性質の階層と理論的特徴

システム理論は探求の対象を“システム”として認識するための認識枠組みを提供している。その枠組みにおいて対象はシステム概念を用いてシステムとして記述される。認識されたシステムが示す性質は理論枠組みの中においてシステムの性質として描写される。

一般にシステムの性質は階層的に捉えることができる[1,4,5]。最下層には入出力システムのプロセス層があり、その上層には意思決定の特性に応じて、制御（管理）層、適応層、自己組織化層がある。各層間の関係はシステムの構造的階層関係を直接表しているとは限らない。たとえば、組織における階層構造との対応をそのまま取れるとはいえない。むしろ、組織の場合でいうならば、Beer[8]のようにシステムの階層を機能的階層の基本単位と捉え、組織を基本単位の再帰的連結関係により表すことも有効である。このシステム論的階層はあくまでシステムの性質の一般的分類と機能的階層関係を表している。このシステムの性質の階層関係を参照して、現在のシステム理論、特に数理的アプローチによるシステム理論における主要な問題を位置付けることができる。

現代制御理論は制御層と適応層の一部における、負のフィードバックによる有効な制御則を効率的に計算するための工学的手法の開発に主眼が置かれている[17]。そこで扱われるシステムモデルは線形モデルがほとんどであり、少なくとも対象が線形近似できるという仮定が必要である。線形モデルによる制御理論については、数理的システム理論の立場

から公理的アプローチによるシステム論的意味付けが展開されている[5]。

最適化を中心とする意思決定問題は、主に制御層において定式化される[5]。Simon の満足化意思決定基準もこのレベルで取り扱われている。組織における下部組織と上部組織の統合問題も制御層における問題として数理的システム理論では一般的に定式化され展開されている[1,4,5]。

このように制御層までのシステムの性質については、その性質を示す重要な基本モデルがいくつか提案され、強力かつ精緻な理論展開を見せている。すなわち、社会・経済システムを制御層までのシステム概念によって考察するならば、工学的な設計問題の社会・経済システムへの含意の妥当性の問題はあるものの、社会・経済システムにおける制御的側面についてはかなりの有効な知見が得られるであろう。

経済システムを Kornai[23]のように実体過程のプロセス層と情報の流れを主体とする制御層の二階層のみで記述することは、塩沢[19]のいうシステム二元論の誤謬に陥る可能性がある。社会・経済システムのような複雑なシステムにシステム理論を適用することは、これまで理論展開が十分な制御層までのモデルのみにより記述することを意味しない。むしろ社会・経済システムにおける適応層、自己組織化層のシステムの性質の解明を主眼とすることが社会・経済システムの本質的理解にとって欠かせない。

しかしながら、適応層、自己組織化層に関するシステム論的基本モデルは発展途上である。たとえば構造をパラメータ化してシステムの自己組織化を扱う試みがある。その本質的な考え方は次のようなものである。まずシステムの構造空間  $A$  を設定し、各システムの構造は構造空間  $A$  上の一点として表現されるようにする。これと時間軸を表現する空間  $T$  を設定し、自己組織化の過程を時間軸空間  $T$  から構造空間  $A$  への写像  $f$  として記述する。自己組織化過程の性質はこの写像  $f$  の数学的性質、たとえば微分方程式系や代数的関係により表されることになる。この方法は何らかの“過程”の数学的記述としてある種の自然さを持っている。後述する遺伝的アルゴリズム (GA) の挙動についても、GA の創始者である Holland は状態概念を用いて構造空間への写像として定式化している[18]。しかしながら、写像は未知の構造への対応をその本質的性質により与えることができない。その意味においてシステムの性質の階層関係における自己組織化層に対する満足な理論展開は未だできていないといえる。システム理論は本来的にそのような複雑システムの問題に第一の関心があったといってもよい。またこの理論展開ができなければ、社会経済システムに対するシステム理論が不十分なことはいうまでもない。

システムの構造を数学的公理によって規定する数理的な一般システム理論の立場からは、自己組織化と構造との関連はゲーデルの不完全性定理により明瞭に位置付けられる[9]。しかしそれにより構造変動のメカニズムまでが明らかになっているわけではない。

### 3. agent-based modeling の特徴

agent-based modeling は、主体の学習とシステムの創発的性質の研究にとって有望であり、演繹的方法、帰納的方法と並ぶ第三の科学的方法あるいは第三の研究モードという見方も存在する[2,20]。この是非はともかく、オートマトンのような従来のシステム理論における典型的モデルでは容易に扱うことができない。オートマトンモデルはシステム理論

においては基本的位置を占めており、オートマトンモデルとの関係を明らかにすることで agent-based modeling の特徴もより鮮明になる。これについては次節で改めて考察する。本節では agent-based modeling の方法論上の特徴を述べる。

agent-based modeling の方法は以下のようなエージェント個人とエージェント集団との意思決定の学習サイクルとして特徴づけることができる[2,3,6,7]。(1)各エージェントは意思決定のためのルールの集合を持っている。(2)そのルールとエージェントの置かれた環境に依存して、次の時点における行動をエージェントごとに意思決定する。(3)意思決定に基づいて各エージェントが行動する。(4)行動の結果、エージェント間、エージェントと環境間の相互作用がエージェントに返ってくる。(5)返ってきた結果と過去の意思決定とを何らかの基準に基づいて比較評価する。(6)評価に基づいてルールの集合を改善する。

agent-based modeling における評価と改善のプロセスでよく使われる手法としては遺伝的アルゴリズム (GA) の方法がある[2,7,18]。遺伝的アルゴリズムは生物進化のメカニズムからヒントを得て作られた一種のシミュレーション技法である。遺伝的アルゴリズムの基礎的な用語には染色体、突然変異等の生物進化に関する用語が使われている。

遺伝的アルゴリズムによるシミュレーションの基本的手順は以下のようになる。

(1) 初期集団を生成し、最初の「世代」とする。各世代は「個体」の集合である。各個体は「染色体」であり、一次元配列として表現される。染色体の各位置(「遺伝子座」)には遺伝情報を持つ「遺伝子」が乗っている。集団の個体数、染色体の長さやコーディング方法は問題ごとに適宜決めなければならない。

(2) 初期集団が生成されると次からは各世代共通の手順が繰り返される。最初に各個体に対し適応度を計算する。すなわち、各個体が状況にどの程度適応しているかを評価する。

(3) 各個体の適応度が計算されたら、その値に基づいて「選択」を行って、交配をする個体対を選ぶ。選択は基本的に適応度の高い個体がより多くの子孫を残せるように行う。その具体的方法はいくつか提案されていて、やはり問題に則して決められる。

(4) 選択された個体対に対し、一定の所与の確率に従って染色体の「交叉」を行う。交叉の方法もいろいろあるが、基本的には双方の染色体の一部を取ってそこから子孫の染色体を作る。

(5) 交叉が終わったら一定の所与の確率に従って「突然変異」を起こさせる。突然変異では染色体の一部の値を変える。

以上の操作が終わると新しい世代の集団ができあがる。この新しい集団に対し再び適応度評価、選択、交叉、突然変異を作用させ次の世代を作ることを繰り返す。

GA の方法は現在もある程度うまくいっており、将来的にも有望な手法といえる。その改善プロセスの中で使われる交叉という遺伝的操作は、染色体でコード化された現世代のルールを組替えて次世代の全く新しいルールを作る。

GA におけるプロセスを記述する数理モデルとしてはマルコフ過程を利用したものがある[15]。選択、交叉、突然変異という GA の遺伝的操作が確率的に個体集団に作用する以上、確率モデルによる記述は自然な流れでもある。しかしながら、ルールの改善過程に対するシステム理論的な興味は、システム構造の時系列的変化パターンの確率的記述という



よりは、構造変化の説明モデルにある。

GA における過程は確率的挙動としてのみでは理解できないというだけでなく、生成されたものが改善されている保証もない。収束先の最適性の判定も困難なことがある。しかし、従来の制御理論と異なり、各主体は目的値を知らなくても学習することができ、全く新しい意思決定構造を作り出す。このような学習方式に関しては、人工知能、認知科学におけるマルチエージェントシステムでの強化学習などが知られている[9,11]。GA が agent-based modeling における有力な方法の一つである理由は、社会・経済システムに対して GA の適応過程が持つ含意にある。GA における各遺伝的操作の社会・経済システムへの直接的類推を行う試みもあるが[15]、システムの進化的学習の原理モデルのプロトタイプとしての役割も重要である。すなわち、上述した agent-based modeling の方法におけるエージェントの意思決定ルールの改善過程を進化的に行う実際のメカニズムを GA は提供してくれる。

### 3. 表現形式としてのオートマトンモデルの適用可能性

オートマトンは状態を持つ入出力システムの基本的モデルである。すなわち「プロセス」はオートマトンにより記述することが基本的に可能である。オートマトンモデルは現代システム理論の一つの範型である。

対象の特徴によりオートマトンから派生した多様なモデルが作られている。たとえば状態遷移が事象（ある時刻に発生する注目するイベント。たとえば「客が到着した」等）により「駆動」するような場合には DEVS(Discrete Event System Specification)[22]、ペトリネット等の離散事象システムモデルがある。あるいは、コンピュータのようにメモリへのアクセスが本質的動作の一部であるような場合のシステム記述に適するチューリングマシンも、モデルとしての歴史的順序はオートマトンの方が後ではあるが、形式的 (formal) 観点からはオートマトンの派生と捉えられる。

オートマトンの形式的定義 (formal definition) は、入力集合、出力集合、状態集合、状態遷移関数、出力関数、初期状態の組で与えられる。各時刻（ただし時間軸もシステムの動作を記述するためにモデル化されている）には入力集合中の特定の要素がシステムに入力され、その時刻に取られていたシステムの状態が次の時刻に状態遷移関数により状態集合中の特定の状態に遷移する。また、入力と状態に依存して次の時刻に出力関数により出力集合中の特定の要素が出力される。ある初期状態を持つ特定のオートマトンは、ある入力列  $x_0, x_1, \dots, x_t, \dots$  が与えられると、状態遷移関数による状態遷移  $s_0, s_1, \dots, s_t, \dots$  と出力関数による出力列  $y_0, y_1, \dots, y_t, \dots$  が決定する。

オートマトンはプロセスレベルのシステムの挙動の一般的記述形式として十分強力である。たとえば制御理論でよく定式化される線形微分方程式系は、入力集合、出力集合、状態集合が線形空間で、状態遷移関数、出力関数が線形微分方程式によって与えられているモデルである。GA の挙動もオートマトンモデルにより定式化することは一応可能である。もっとも素直な定式化は全染色体の膨大な状態集合にオペレータを入力とする状態遷移関数を定義する。この定式化はオートマトンモデルのもつ記述力の一般性により可能となったものである。われわれが理解したいのは前節で述べたように、各エージェントの意思決

定ルールの学習方法である。オートマトンモデルによる GA の“素直な”定式化はその方法を記述していない。しかし、学習方法そのものをオートマトンモデルで記述することは困難である。

オートマトンモデルを社会システム、経済システムの記述に利用する有力な方法のひとつは、社会・経済システムを巨大なオートマトンのネットワークとしてモデル化することである。たとえばシステムの最小単位をひとつのオートマトンとして記述し、組織、機関、社会等をそれらの結合として表現する。結合された複合オートマトンは上述の形式的定義を満たすオートマトンとして記述できる。結合された複合オートマトン同士をさらに結合することも可能である。塩沢[19]は最小単位としてシステムを構成する人間をオートマトンとして捉え、そのネットワークとして経済システムを表現することを提案している。この場合主体の定型的、習慣的行動が単位的オートマトンにより表現されるといえる[19]。

オートマトンの行動を制御するシステムをオートマトンにより与えることも可能である。工学的にはオートマトンの制御問題として詳細な研究が発展している。Kornai[23]は経済システムを実物過程と制御過程とに分離し、経済システムの二元論的記述をオートマトンモデルの形式に基づいて行っている。

オートマトンモデルにより表現されたシステムの環境変動は入力集合の変動として記述されるのが自然である。すなわち環境が変動することにより、入力列  $a_0, a_1, \dots, a_t, \dots$  から入力列  $b_0, b_1, \dots, b_t, \dots$  に入力に変化するか、あるいは入力集合の要素（入力アルファベット）が変化する。このときシステムが適応する方法としてはいくつか考えられる。適応層のレベルでは、入力列の変化に対応可能な制御則を与える。ただし、入力列の変動が“予想”を大きく越える場合や、入力アルファベットそのものが変化するほどの大きな環境変動では適応層レベルでは対応できないのが普通である。自己組織化層においては、サブシステム間の結合様式が変化したり、状態遷移関数や出力関数といった各サブシステムの構造が変化する。これらにより全体システム自体の構造を変化する。

社会・経済システムの時間的変動過程を“ひとつ”の複合オートマトンの挙動として理解する自然な方法は、システムの変動をオートマトンの状態の推移として扱うことである。しかし、かなり大胆なマクロ的視点から定式化しない限り、過去の社会・経済システム歴史的変動ですらそれを説明するオートマトンモデルの記述は困難であろう。もし単位的オートマトンを各人の定型的行動のレベルにとるならば、社会全体を表現する“ひとつ”の複合オートマトンの状態推移としてそのような社会・経済システムの大きな変動を再現するのはほとんど見込みがない。そうではなく、環境変動とシステム変動の相互作用として上述した複合オートマトンの自己組織化の観点から扱う方が実り多い。この過程を“別の”オートマトンの状態推移として記述しようとするのは、今述べた隘路に落ちてしまう。

オートマトンモデルはシステムの記述形式として非常に有効であるが、それを社会・経済システムの記述に適用するときには、塩沢の提案するオートマトンネットワークのような複合オートマトンを基本枠組みとして、複合オートマトンそのものの構造変動を対象とした全体システムの変動の解析を行う必要があるだろう。agent-based modeling はそのための方法を提供してくれる。GA は具体的な解析ツールのひとつであり、変動のモデルである。

#### 4. 意思決定ルールの改善のための“説明”としてのシステムモデルの必要性

agent-based modeling における各エージェントの持つ意思決定ルールの改善プロセスのシステム理論的探求はシステムの自己組織化過程の解明に本質的役割を果たすことが期待できる。そのためにはまず意思決定ルールの改善プロセスを記述するシステムモデルを構築する必要がある。

ある注目する事象または対象に対してその関心領域を定め、事象を構成する要素とその間の相互作用を規定してシステムとして認識する。そのシステムは知的構築物であり、必ずしも間主観性を持った表現を与えられていない。それを表現したものがシステム理論におけるシステムモデルであり、注目する事象の性質の理解を与えるものである。数理的システム理論においては、システムモデルの記述に“数理的表現”を用いる。特に代数、位相、論理など主に今世紀に発達した定性的数学は、数学的对象の関係性に基いた数学的構造を表現するものであり、それらを用いれば、例えば線形システムの記述に微分方程式系を用いるとき必要な実数体の性質等をアプリオリに仮定せずに、システムの相互作用の一般的記述を直接的に行える。また論理学における公理的表現を利用することで、システムの構造を一般的に規定することが可能で、同型な構造を持つシステムのクラスの議論が厳密に行える[1]。

ここで注意すべきは表現されたシステムモデルと表現したい対象との関係である。システムモデルは対象の“すべて”を表現しているのではなく、関心領域に従って切り取られた対象のある“領域”を表現している。agent-based modeling では、注目する事象は通常シミュレーションによって得られたシステムの挙動である。一般的にこれらの挙動は複雑である。挙動が含意する社会的意味と説明を与えるのがシステムモデルである。シミュレーションモデルは複雑な挙動を生成するためのシステムの構造を与えてはいるが、説明を与えているとは限らない。関心領域に依存して切り取られる挙動の集合は異なり、複数のシステムモデルの複合が complex system においては不可欠である。単一のシステムモデルですべての挙動を説明しようとすることはシステム理論におけるシステムモデルの構成過程からは意図されない。また、モデルについてはよく「モデルは対象の単純化である」といわれる。しかし“説明”のためのシステムモデルは対象の単なる単純化ではない。“説明”のためのシステムモデルは、注目する事象の必要十分な説明を与える表現であり、その意味で対象の関心領域に関する理解が深まるものである。

#### 4. ルールの改善のためのシステム理論の課題

agent-based modeling における意思決定の学習サイクルでは、各エージェントは、意思決定の前時点での結果を現時点における意思決定の入力のひとつとしているという意味で、フィードバックによる学習を行っていると言える。しかし制御層におけるシステム理論におけるフィードバック理論と本質的に異なる点は、意思決定のルールの改善が関数的でないことである。すなわち、改善前のルール集合  $A$  と改善後のルール集合  $B$ 、そして改善過程  $f$  によって、 $f:A \rightarrow B$  という数学的構造として把握することが困難である。GA のように遺伝的操作がある確率分布に従って適用されるならば、 $f$  をある種の確率遷移モデルとし

て記述することも可能である。実際多くの数理モデルはマルコフ過程を基本にしている。これは確率構造をエージェント集団からなるシステムにおける挙動と生み出す“深層構造”として捉えていることになる。それは個々のエージェントの意思決定とその相互作用から集団としての進化的学習という agent-based modeling における中心的な関心にある問題に必ずしも満足 of いく説明を与えるものではない。

では確率モデル以外にエージェントの進化的学習を説明する数理的システムモデルはあるのだろうか。これが agent-based modeling の理論化に対し悲観的な点である。少なくとも現在の数学における「関数」あるいは「関係」の定義からは、 $f:A \rightarrow B$  の関係を直接的に表現することはできそうにない。「関数」の場合、ドメイン集合  $A$  とコドメイン集合  $B$  を確定し、集合  $A$  の“すべて”の要素に対して唯一の集合  $B$  の要素を対応させることが関数の定義十分性 (well-defined) である。コドメインは学習先の集合であり、これが前もって確定していることは一種の論点先取である。

ドメインやコドメインが確定していてもそれらが十分に大きく、対応関係も個々のドメインの要素に対する計算規則ともいえるべきものが与えられていて、毎回計算する必要があるならば、あらかじめ学習先が所与とは考えられない、という見方も存在するだろう。

例えば、ドメインが巨大な探索空間であり、その中から最適値を探索する山登り法あるいはランダム法等のアルゴリズムが  $f$  の計算規則として与えられているような場合が典型である。このときには確かに“行き先”である最適値は所与ではない。それは探索される対象である。そして関数  $f$  は例えば  $\max\{g(x) \mid x \text{ は } A \text{ の要素}\}$  と表現される。しかし、これは“探索問題”の表現であり、最適値に到達されるまでのプロセスを与えてはいない。そのプロセスは探索アルゴリズムによって与えられる。アルゴリズムは次にどちらの方向を探索するかという局所的な探索ルールを示しているに過ぎない。このアルゴリズムの性質、すなわち探索プロセスの性質こそが進化的学習における問題関心に相当するものであり、これは上の表現からは自明とはいえない。

遺伝的操作の交叉等を「関数」として“操作的”に定義して進化的改善プロセスを表現することも局所的で段階的な手続きを与えるに過ぎない。特定の対応関係は事後的にしか確定できない。交叉確率等による確率構造を入れる方法は前述の隘路に陥ってしまう。

上述のことがらは、 $f$  の部分をどのように構造化したとしても、関数概念の枠組みで考える限り困難であることを示している。前述したオートマトンモデルも関数概念の枠組みで記述されるモデルの典型例であり、GA あるいは意思決定ルールの進化過程の説明モデルとしては不十分である。

意思決定のルールの改善プロセスは関数関係にならなくとも、現在の数理的システム理論の枠組みで改善プロセスを関数的に扱うことは困難である。しかしわれわれの目標は、遺伝的アルゴリズムあるいは交叉、突然変異等の遺伝的操作に関する数学的理論を作ることではない。交叉と同じ効果を生むような過程を含んだモデルを記述することで改善プロセスを表現することである。

## 5. 社会・経済システム研究のための agent-based modeling のシステム理論の方向

システム理論はシステム全体とシステムを構成する要素の相互作用を扱うための枠組み

の定式化の方法を与えてくれる。社会や経済は構成要素もシステムであるような複合システム (complex system) として認識されるのが普通である。そのときの主要な相互作用は上位システムと下位システムとの関係になる。上位・下位の概念は相対的である。すなわち自己を含むシステムに対しては下位となり、自己が含むシステムに対しては上位になる。このような捉え方はシステム運動の当初より行われてきた[21]。しかしながら理論的、数理的には制御層までの理論が大きな発展をしてきた。

agent-based modeling は適応層、自己組織化層におけるシステムの性質を扱う方法を与えてくれるが、その本質は各エージェントの行動ルールの変更過程とそれが上位システムとの相互作用により規定される点にある。すなわち各エージェントが各自の行動ルールをもち、それを自律的に変更し、その変更の過程には各エージェントの上位システムとの相互作用が本質的役割を果たす。このことは、たとえばエージェントの行動あるいは意思決定のルールをゲーム形式により記述したとしても、全体のシステムの性質は通常のゲーム理論の方法では扱えないことを意味する。通常のゲーム理論においては各自の行動ルールは共通知識 (common knowledge) として共有され、その変更や上位システムとの相互作用は理論の埒外である。いわゆる進化ゲーム[26]も記述の形式においては通常のゲーム理論の前提を踏襲している限り限界がある[20,24]。

各エージェントの行動ルールをゲーム形式 (プレーヤ集合, 戦略集合, 利得関数) で記述するとき, agent-based modeling では各自がまったく独立に行動ルールを認識し, エージェント間で共通知識とはなっていない。各エージェントの行動ルールに基づいてゲームのプレイが実行され, 各エージェント間の相互作用が生じる。プレイの結果に依存して, 各エージェントはその行動ルールを改善する。各エージェントの行動ルールを独立にゲーム形式で認識する典型的な枠組みはハイパーゲームと呼ばれ, その改善過程に GA を適用することができる[24,25]。

上述のような特徴をもつ agent-based modeling は社会・経済システムのどのようなシステムの性質を明らかにしてくれるのであろうか。いくつか考えられるであろうが, もっとも有望な問題のひとつにミクロ・マクロループに関する問題がある。塩沢[19]はミクロ・マクロループの問題は社会・経済システムにおけるシステム二元論を超えるものと指摘している。agent-based modeling の方法は, 各エージェントの行動ルールの学習のミクロループとシステム全体の学習とエージェントへの相互作用というマクロループを扱うことができる。

具体的な問題領域の例として, 組織論における組織学習の問題がある。組織学習は組織の学習曲線のような概念により記述されることが多い[27]。しかしシステム概念 (たとえばサイバネティクス等) の援用により, Espejo[13]らは組織学習のモデルを提出している。組織の各構成員の学習であるシングルループと組織全体の学習であるダブルループからなるモデルである。そのモデルは agent-based modeling の方法で今後発展が望まれる。

## 6. おわりに

システム理論の一つの重要な目的は, システムの性質の透明な理解がある[1,5]。シミュレーションは改善過程の現象を観察はできるが, その透明な構造的理解のためにはシステ

ム理論化する必要がある。そのためにはシミュレーションによる観察とその知見は必須である。これらはシステムの自己組織化層の理解のための両輪である。本論で考察したような困難性はあるが、システムモデルの意義と必要性を見極めれば agent-based modeling における理論の役割に対する悲観的な見方ばかりではなくなるだろう。

システム理論はある特定のモデルについての理論ではなく、探求の対象の理解のためのモデルは対象領域と理解のレベルに応じて開発されてきた。これまで工学的観点から発達した制御系のモデルが社会・経済分野を含め多様な領域に適用され、有効な一定の役割を果たしてきたことは確かである。しかし、社会・経済システムの生来的複雑性の理解にとりそれらが十分かどうかは疑問がある。既存のシステムモデルの適用で理解が困難な複雑性については新たなアプローチによるシステムモデルを開発していく必要がある。これまではシステム理論の数理的発展は主に工学的な対象領域において進められてきた部分がある。今後は、社会・経済の対象領域におけるシステムの生来的複雑性の理解のためのシステム理論の発展が望まれる。

#### 参考文献

- 1.Takahashi,S. & Takahara,Y., Logical Approach to Systems Theory, Springer-Verlag,1995.
- 2.Axelrod,R., The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration, Princeton University Press,1997.
- 3.Hanappi,H., Evolutionary Economics, Avebury,1994.
- 4.Mesarovic,M.D.,Macko,D. & Takahara,Y.,Theory of Hierarchical, Multilevel Systems, NY:Academic Press,1975.
- 5.Mesarovic,M.D. & Takahara,Y.,Abstract Systems Theory, Springer,1989.
- 6.Takahashi,S.,Evolutionary Learning in Agent-based Modeling,International Conference on AI,Simulation and Modeling,2000(to appear)
- 7.Takahashi,S.,Hinago,T.,Inohara,T. & Nakano,B., Evolutionary Approach to Three-person Hypergame Situation, IEEE International Conference on Systems, Man and Cybernetics, 1999.
- 8.S.Beer, Brain of the Firm, 2nd ed., Chichester, 1981.
- 9.G.Weiss(ed.), Multiagent Systems-A Modern Approach to Distributed Artificial Intelligence, MIT Press, 1999.
- 10.高橋真吾,システムの構造変動に対してゲーデルの不完全性定理の持つ意味, 第17回計測自動制御学会システムシンポジウム講演論文集 pp.443-450,1991
- 11.山田誠二,適応エージェント,共立出版,1997
- 12.R.J.Gaylord & L.J.D'Andria, Simulating Society-A Mathematica Toolkit for Modeling Socioeconomic Behavior, Springer, 1998.
- 13.R.Espejo,W.Schuhmann,M.Schwaninger & U.Bilello, Organizational Transformation and Learning-A Cybernetic Approach to Management, Wiley, 1996.
- 14.M.J.Prietula,K.M.Carley & L.Gasser(eds.), Simulating Organizations, MIT Press, 1998.

- 15.H.Dawid, Adaptive Learning by Genetic Algorithms·Analytical Results and Applications to Economic Models, Springer, 1996, 1999(2nd ed.).
- 16.P.Checkland, Systems Thinking, Systems Practice, Wiley, 1981.
- 17.高橋安人, システムと制御,岩波書店,1983.
- 18.J.H.Holland, Adaptation in Natural and Artificial Systems,MIT Press, 1992.
- 19.塩沢由典,複雑さの帰結,NTT 出版,1997.
- 20.塩沢由典編,方法としての進化,進化経済学会,2000.
- 21.T.D.Bowler, General Systems Thinking, Its Scope and Applicability, Elsevier North Holland, 1981. (T.ダウニング・パウラー,応用一般システム思考,中野文平訳,紀伊国屋書店,1983)
- 22.B.Zeigler,H.Praehofer and T.G.Kim, Theory of Modeling and Simulation, Second Edition, 2000.
- 23.J.Kornai, Anti-Equilibrium, Amsterdam: North-Holland, 1971. (J.Kornai,岩城博司・岩城淳子訳,反均衡の経済学,日本経済新聞社,1975)
24. 高橋真吾, ハイパーゲーム型社会状況における学習: コミュニケーションによる情報生成, 進化経済学会論集第2集, pp. 314-322, 1998.
25. 木嶋恭一編, マルチメディア社会システムの諸相, 1997.
26. J.W.Weibull, Evolutionary Game Theory, MIT Press,1996.
- 27.森本三男,経営組織論,放送大学教育振興会,1987.

## Hard Problems in Mathematical Systems Theory for Agent-based Modeling

Shingo Takahashi

Chiba Institute of Technology

**Abstract** Agent-based modeling shows a fruitful approach to understanding self-organizing behavior in complex social systems. The problems on self-organization have been the primary concern of systems theory. Although there have been a lot of researches studying self-organization, we would need some new way to describe a model for emergent properties and learning process. Agent-based modeling can be expected to provide some novel concept and method on learning process in complex social systems.

The essential features of agent-based modeling include a learning process of the rules based on which each agent makes a decision for his next action. Genetic algorithm is a powerful tool to implement the learning process in an evolutionary manner.

This paper first provides the meaning and role of system models to describe the learning process in self-organization layer. Then the essential difficulties to develop such system models are discussed. The core step in the learning process in agent-based modeling is to produce a new rule set from an old one in the previous learning phase. This correspondence is very hard to describe as some “function”

between the rule sets. We should develop more effective mathematical device and concept to build system models of the learning process.



システム理論からの組織論へのアプローチ  
ー組織学習の進化的システムモデルの枠組みー  
An Approach to Organization Theory from Systems Theory  
-A Framework of Evolutionary System Models for Organizational Learning-

高橋 真吾

要旨

システム理論の立場から組織論へアプローチにとり必要なことは、説明のためのエージェントベースのシステムモデルの構築である。本論では組織論の問題として組織学習を取り上げ、そこでの double-loop learning を有効に行うための進化的方法の枠組みを提示した。その核心は各エージェントの持つ内部モデルの学習で、遺伝的アルゴリズムによるシミュレーションがシステムモデルの開発にとって重要な位置を占める。

英文要旨

An approach to organization theory based on systems theory should essentially include the way of constructing agent-based system models to explain the mechanisms how systems properties of interest are produced.

This paper particularly shows a framework of organizational learning by evolutionary mannar to perform effectively double-loop learning. The core step of the learning is described as the learning process of internal models of agents. Also this paper develops a simulation model to be performed by genetic algorithm for learning the internal models of agents. Finally this paper points out that agent-based simulation can help us to construct system models for organizational learning.

(102 words)

キーワード

システム理論, 組織論, 組織学習, 遺伝的アルゴリズム

## 1. はじめに

組織はシステムとして認識できる。システムとはもっとも広義には、認識の対象を構成する要素間のある特定の認識関心に従った相互作用から成る概念的構築物である。システム認識は様々な領域において現れるので、システムの定義自体多様である。「要素間の相互作用」はそれらの共通項であり、どのようなシステムの定義にも含まれるシステムの本質的定義である。システム科学はシステム哲学、一般システム理論、システム理論各論、システム方法論、システム工学等を包含し、システム認識のための概念を基礎に展開される大きな知的領域を形成している。システム理論は同型性によって特徴付けられたシステムのクラスに関する性質を論理的に導く知的企てである。「組織」はシステム科学あるいはシステム理論の対象として興味深くまた面白い。その面白さは組織のシステムとしての複雑性に由来する。

システム理論による組織論の新たな展開を行う際には、組織の存続、成長といった問題に対する議論の有効性が一つの鍵になる。これらはシステムの自己組織性や創発性の問題と密接に関連し、組織論への新たな展開は自己組織性の新たな展開をも意味している。

Barnard(1938)の協働システムとしての(公式)組織の定義が現れて以来、組織をシステムアプローチによって統一的にとらえることが行われてきた。Beer(1981)のサイバネティックアプローチは組織システムモデルのひとつの範型である。組織は複雑であるがゆえにそれを記述するシステムモデルも多様である。

近年における Carley and Gasser(1999)を中心とした Computational and Mathematical Organization Theory は、プロセスとしての組織行動をブラックボックスとするのではなく、組織としての創発的行動を生み出す“メカニズム”を記述するモデルを明示化する試みのひとつである。

Espejo, Schuhmann, Schwaninger and Bilello(1996)は double-loop learning の概念に基づく組織学習のモデルを提示している。組織学習の本質的課題はその有効な学習メカニズムの解明にある。しかし、どのようなメカニズムで double-loop learning を行えば有効であるかについて十分な知見は得られていない。

本論では、システム科学とくに数理的なシステム理論の立場から、組織というシステムにアプローチする方法と、現在までのシステム理論の成果と今後の方向について記述する。特に組織論における問題領域のひとつである組織学習について、エージェントベースの考え方に基づく進化的方法の枠組みとそのシミュレーションモデルを提示し、有効な組織学習のメカニズムを理解するためのシステムモデル構築に向けた可能性を考える。

## 2. 説明としてのシステムモデル

システム理論からのアプローチといっても唯一の手順と呼べるものがあるわけではない。

システムアプローチも、システム工学における最適意思決定のようなハードな方法から、複数の問題関与者を前提としたソフトな方法論まで多種多様である。しかし、どのシステムアプローチにも共通に含まれる本質的要素は“システムモデル”を構築することにある。

組織のような複雑な対象を見るために多様な認識関心があるだろう。日常的業務プロセスに関わる側面、経営戦略的側面、組織内の人間関係の側面、財務的側面等様々である。組織全体は、特定の認識関心により捉えられた多数のサブシステム間の相互作用を持つ複合的な複雑システム（complex system）と考えることができる。

組織のような対象を構成する要素と要素間の相互作用関係にある認識関心から認識した概念的構築物がシステムである。システムモデルは認識されたシステムの表現である。モデルは対象の単純化であると言われるが、本論でのシステムモデルの役割は対象の単なる単純化としてのものではない。システムモデルは認識者の認識関心に従った対象の持つある側面の関係を表現したものであり、対象の“すべて”を含んではいないしシステム理論としてはその必要もない。その意味では“単純化”と言えるだろう。しかしながら、システムモデルには重要な役割がある。システムモデルとして対象の相互作用関係が表現されることで、注目する対象のシステムとしての性質の説明が得られ、対象に対するわれわれの理解がなされる。すなわち説明のためのシステムモデルは単なる対象の単純化ではない。

組織のような複雑な対象に対するシステム理論からのアプローチは、組織に関するわれわれが認識する問題に対する“説明”を与えるものとしてのシステムモデルを構築することである。これが本論での立場である。決して組織を“単純化”して見てみようというのではない。

本論でのシステム理論からのアプローチでは、認識されたシステムの表現であるシステムモデルを数理的に記述することを意図している。特に20世紀において発展した集合論、代数、数理論理学を含む定性的数学の形式は関係の記述を基礎としており、システムモデルの表現にとり有益というに留まらず本質的でさえある。しかしながら、複雑システムの表現は、自己組織化等のシステムの性質を扱わねばならず、現在まで必ずしも十分とは言えないし、今後も困難な課題が山積していることも事実である(高橋,2000b)。

しかしそのような本質的課題の存在にもかかわらず、数理的な表現を利用するアプローチは、学術的な議論を行いたいということでは決してなく、組織の説明、理解のためのシステムモデル構築とともに、むしろこれからのIT革命時代における実践的な理論をも目指している。環境が激変する時代における組織の変革は、各構成員の能力をいかに組織としてシステム化するかにかかっており、そのためには暗黙的な了解事項や知識を組織として共有化することが不可欠である。また、そのための手段として情報システムが本質的構成要素として含まれることが必要で、それは共有知識の形式化の必然性を意味する。すなわち、これからのIT革命時代においては、組織変革のための数理的に形式化された組織の理論がその実践のために必要である。

### 3. 組織の機能階層

組織のシステムとしての複雑性には大きく分けて二つの特徴がある。一つは複数のサブシステムの多重的な結合として全体のシステムが成り立っている。これは複雑システム (complex system) の基本的記述形式ともなっている。結合の多重性はたとえば Beer の組織モデルにあるような、機能の入れ子構造により表現されることもある。複雑性の他の特徴は、認識の多相性である。認識される状況は多様な認識関心からシステムとして認識される。システムモデルは対象の認識関心に従った表現である。すなわち組織は“単一の”多重的結合システムモデルにより表現することは困難である。

組織全体を俯瞰するための参照モデルとしてサイバネティックモデルに基づく組織の機能階層を数理的組織論の立場から表現した高原の組織モデルがある (図 1)(Takahara,1999)。まずこのモデルについていくつか注意しておくことがある。このモデルは組織というシステムの“機能”に焦点を当てた階層を表現しており、権限関係に基づくいわゆる組織の階層構造を表現したものではない。また、各レベルは複雑なサブシステムを形成しており、それらについてはこのモデルでは表現されていない。それらはこのモデルを出発点とする数理的組織論の中で描写されていく。階層間についても次節以降で述べるようなマクロ-ミクロループの形成など今後このモデル上に明示的に位置付けていかなければならない問題もある。このモデルは組織という複雑な現実をあまりに単純化していると見る人もあるかもしれない。しかしこれは組織のシステムの性質の機能としてもっとも大きな意味のカテゴリを階層という形式で表現したものであり、組織のシステム理論的問題の所在の相互関連の明確化のための参照モデルとして有効である。逆にいえばこのモデルのみを見ていても組織が理解できるのではなく、当該の問題がこのモデルのどこに位置するかを把握し、それに関連するサブシステムあるいはサブシステム間の相互作用の複雑な過程を考えていくための、いわば出発点をこのモデルは提供している。

このモデルでは組織は四つの機能階層から構成される。最下層は process レベルで各業務に対応する。各プロセスはある目標に向かって進められ、またプロセスは相互作用をしている。ここまではモデルにおける operational レベルとして表現される。プロセスをどのように統合し全体の仕事を最適化していくかが管理の問題となる。organizational レベルではこの統合の問題を扱う。institutional レベルでは環境（組織の置かれた状況）を観測し、組織全体のマクロ的目標を定め、組織の存続と発展のための意思決定を行う。

このモデルは一般的な“見取り図”であり、各レベルはより具体的詳細に記述されなければならない。現在のところ数理的システム理論として process レベルと operational レベルに関しては精緻な議論が展開されている。organizational レベルでは階層システム理論における統合理論が援用され、相互干渉予測原理等の統合原則が定式化され有効な理論が構築されている (Mesarovic, Macko and Takahara,1975; Mesarovic and Takahara,1989)。

institutional レベルについても、最小多様度の法則とその定式化等システム理論の様々な知見が知られており、極めて豊富な議論が展開されている。

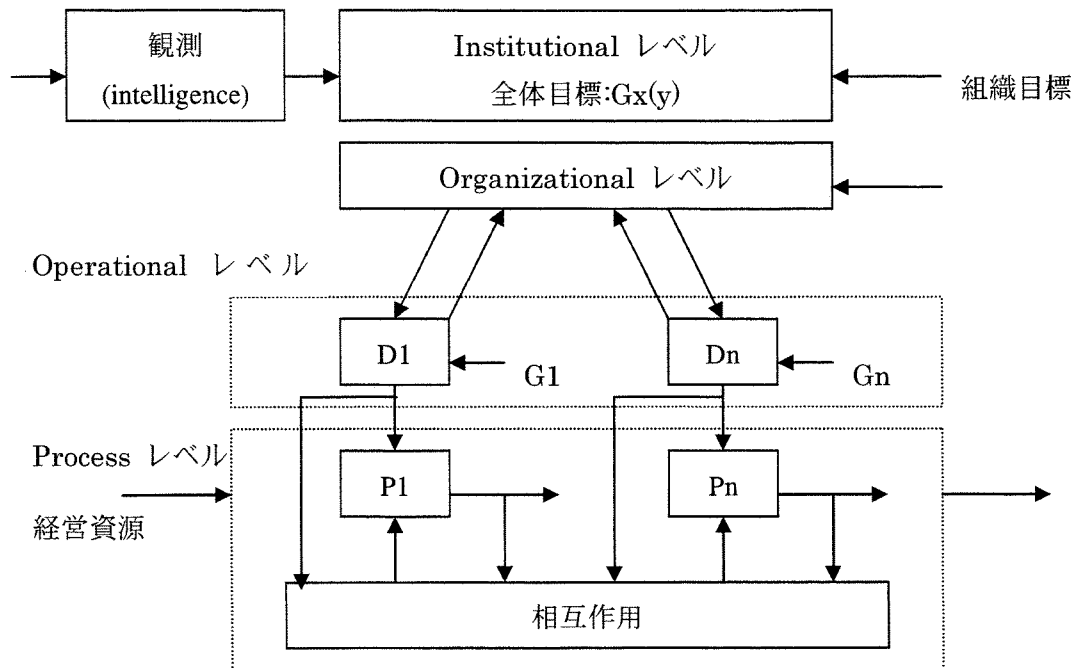


図1 組織の基本的サイバネティックモデル

#### 4. ミクロマクロループの形成過程と組織学習

組織論におけるひとつの論点として、ミクロマクロループの形成に関する問題がある。現代の激変する環境に置かれた企業組織では、経済活動のすべてを市場に任せたり、あるいは反対に組織へと内部化したりするだけでは組織の成長だけでなく組織としての存続そのものが困難になっている。複雑な環境への適応に対して、“ネットワーク”という概念の有効性が以前から唱えられている(今井,金子,1988;出口,1994;塩沢,1997)。ネットワークの構築には、各構成員の自律的行動から組織全体へのリンクであるミクロマクロループ形成の過程が本質的であることも指摘されている。

ミクロマクロループ形成の問題に関連した重要な論点として組織学習の問題がある。組織学習は、組織の行動に有効性を与える組織の能力の向上と定義できる(Espejo et al,1996)。institutional レベルでは環境の多様度と組織の多様度との調和を目指すことが一つの目標である。ではどのように環境または組織の多様度を有効に調節したらよいか。組織の学習は構成員個人の学習と本質的に連関している。個人の学習なくしては組織の学習はありえないし、組織の学習により行われる組織の行動が個人の行動や意思決定の制約条件ともなる。すなわち、組織学習の問題はミクロマクロループの形成問題の典型例であ

る。

この場合、ミクロループは single-loop learning と呼ばれ、構成員各自の学習であり、ミクロマクロループの形成は double-loop learning と呼ばれ、構成員の組織環境に対する解釈枠組み（メンタルモデル、内部モデル）の学習に相当する(Espejo et al,1996)。double-loop learning の本質は内部モデルの学習にある。そのとき double-loop learning を個人の内部モデル学習と組織による内部モデルの共有化とを考えることができる。Espejo et al(1996)は前者を individual double-loop learning、後者を organizational double-loop learning と呼んで区別している。

「数理的組織論」に組織学習の位置をマッピングすることは可能である。構成員各自の学習である single-loop learning は「組織の全体モデル」の各レベルにおける個々の意思決定問題に相当する。double-loop learning はたとえば process レベルにおける各自の内部モデルの学習から institutional レベルにおける内部モデルの共有化へとリンクする部分で生じる学習と考えることができる。

#### 4. 組織における内部モデル原理

組織のような複雑な社会システムは意思決定が相互に関連する複数のエージェントから成り立っている。各エージェントが他のエージェントがどのように意思決定し行動するかについての特徴を認識することはエージェントの意思決定にとり不可欠である。この認識は状況の内部モデルとして各エージェントが構築する。内部モデルは複数のエージェント間の相互作用を含む複雑な状況のエージェント自身が持つモデルである。内部モデルはエージェントの行動を記述する内部に構成される。

たとえば状況をゲームの形式により記述するような場合には、内部モデルはエージェントの戦略や戦略間の選好等のゲームのルールを表現している。内部モデルは各エージェントが意思決定の際に参照される。たとえば2人のエージェントが含まれる状況において、一方のエージェントが代替案 x より y の方を選好し、他方のエージェントは y より x を選好するとこのエージェントが信じているならば、このエージェントは代替案 x を選択する。しかし、他方のエージェントが実際は x より y を選好するならば両者にはコンフリクトが発生する。

内部モデルの概念は元々制御理論において導入された。その主要な結果は次のような内部モデル原理として知られている(Connant and Ashby,1970)。・“a control system should include a regulator isomorphic to the system to be controlled”・システムの環境が制御の対象であってもこの内部モデル原理は適用できる。

制御理論における内部モデル原理は、制御者と被制御者という二元的枠組みの中で有効であり、組織にはそのままでは適用できない。しかし、エージェントとその内部モデルと

いう概念枠組みで組織を考える場合、組織学習の過程を説明、理解するためのシステムモデルの基礎として“組織における”内部モデル原理を考えることができる。

組織における内部モデル原理として次のようなことが考えられる。エージェントは“良い”意思決定を行うためには複雑な状況に整合した内部モデルを構築しなければならない。これは“予想 (conjecture)”である。

## 5. 組織学習のシステム理論からの基本的枠組み

本節ではエージェントベースの観点から組織学習を構成する基本的な枠組みを記述する。エージェントは最小のプロセス実行単位である。図 1 で示したモデルの各機能階層はエージェントの行う行動により実現される。すなわち各機能階層は相互作用するエージェントのネットワークにより実現される。したがって、エージェントのあるネットワークが図 1 における四つの機能階層をすべて実現していると考えてもよい。また、トップマネジメントから業務処理までの組織階層との関係と図 1 の一般モデルとを対比させるならば、Beer(1981)の組織のサイバネティックモデルのように、各組織階層が四つの機能階層を持つ入れ子の構造をしていると考えることもできる。

本節で提示する組織学習のための基本的枠組みは以下になる。

- (1) エージェントはネットワークを形成している。
- (2) 各エージェントは状況の内部モデルを持つ。
- (3) 各エージェントは自己の内部モデルを参照して意思決定し行動する。
- (4) 自己の行動の結果と他のエージェントの行動は状況からのレスポンスとして観察され受け取る。
- (5) 観察し受け取ったレスポンスに基づき各エージェントは自己の内部モデルを修正する。

この枠組みの上で、組織学習の過程は次のように行われる。

### (1) individual single-loop learning

各エージェントが自己の内部モデルを参照して意思決定する過程である。エージェントが合理的意思決定をすると仮定するときは、individual single-loop learning は最適化の過程と一致する。

### (2) individual double-loop learning

各エージェントが状況からのレスポンスに基づいて自己の内部モデルを修正する過程である。

### (3) organizational single-loop learning

各エージェントの行動の集合による組織としての行動である。

### (4) organizational double-loop learning

各エージェントの内部モデルの共有化の過程である。

上記の組織学習の過程で最も問題となるのが二つの double-loop learning である。

Individual double-loop learning では各エージェントが独立に自己の内部モデルを修正する。しかし、中立的な“正しい”モデルが存在してそれとの差を縮小するといった通常のコントロールの概念に沿った方法で修正をすることはできない。すなわち状況の“正しい”モデルを知らずに各エージェントは自己の内部モデルを修正しなければならない。さらに、修正された内部モデルが“正しい”かどうかの判断も基本的にはできない。

organizational double-loop learning はさらに困難な過程のように見える。エージェントのネットワークにおいて内部モデルが共有されるとは、最も狭い意味では、各エージェントの持つ内部モデルが一致することである。各エージェントは自己の内部モデルを明示的に他のエージェントに示すことはしない。したがってエージェントのネットワークにおいてある一つのモデルが共有されたかどうかを各エージェントが個別に判断することは、内部モデルの比較という方法ではできない。内部モデルの共有化は各エージェントの内部モデルの修正の結果として得られる。すなわち有効な organizational double-loop learning の方法は、内部モデルの共有化を導くような individual double-loop learning、内部モデルの修正方法ということになる。

## 6. double-loop learning（内部モデルの学習）と進化的方法

### 6. 1 基本的ステップ

内部モデルの学習ということは、内部モデルの“変化”を意味する。これは、複数の代替案から最適なモデルを選択するような最適化問題を解くことでもないし、目標となる理想モデル目指して漸近的に近づく学習プロセスでもない。事後的にモデルを評価して、その評価値を次の変化ステップでの変化トリガーにするような、いわば“進化的”なプロセスである。

本論で提示する有効な organizational double-loop learning のための方法は進化的方法に基づくものである。この基本的なステップは以下である。

- (1) 各エージェントは意思決定のためのルールを持っている。
- (2) そのルールとエージェントの置かれた環境に依存して、次の時点における行動をエージェントごとに意思決定する。
- (3) 意思決定に基づいて各エージェントが行動する。
- (4) 行動の結果、エージェント間、エージェントと環境間の相互作用がエージェントに返ってくる。
- (5) 返ってきた結果と過去の意思決定とを何らかの基準に基づいて比較評価する。
- (6) 評価に基づいてルールの集合を改善する。

ここではルールという形で各エージェントの内部モデルが記述されていると考えてよい。



第2ステップの各エージェントの意思決定が individual single-loop learning であり, 第6ステップのルール集合の改善が individual double-loop learning である.

ルールの改善プロセスについてのシステム理論的探求は緒についたばかりである. 改善プロセスを記述し説明するシステムモデルが構築できていない. しかしながら, 改善プロセスを計算機シミュレーションにより実現する試みは精力的に進められている. マルチエージェントシステムにおける分散人工知能等のエージェント技術も有効に利用される. その有力な手法の一つが遺伝的アルゴリズム (GA) の方法である (Axelrod,1997;Takahashi,Hinago,Inohara,and Nakano,1999;Holland,1992). 遺伝的アルゴリズムは生物進化のメカニズムからヒントを得て作られた一種のシミュレーション技法である. organizational double-loop learning は事後の評価と改善が繰り返され, エージェント間で内部モデルが共有されることである.

遺伝的アルゴリズムの適用は次の意味で自然であり重要である. それは organizational double-loop learning のための進化的方法におけるエージェント間の情報交換と内部モデルの改善を実行するための具体的計算アルゴリズムを与えている. あるいは遺伝的アルゴリズムの枠組みそのものがそれらに対するひとつのモデルであるとも考えることもできる. さらに遺伝的アルゴリズムにおける世代の発展を内部モデルが学習される過程になぞらえることができる. すなわち状況認知が改善され, 内部モデルが状況に適応していく様をある意味で直接観察することができる.

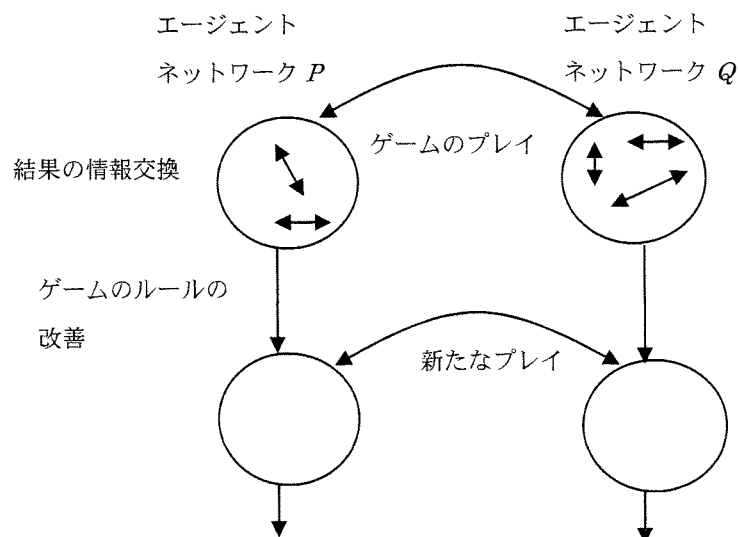


図2 ネットワーク型ハイパーゲームにおける基本的改善プロセス

## 6. 2 シミュレーションによる実験

遺伝的アルゴリズムにより実際にシミュレーションを行うには, エージェントの内部モ

デルを表現する形式を定める必要がある。これまでの研究ではハイパーゲームと呼ばれる形式と資源分配モデルに基づいた記述による内部モデルの表現が主に扱われている。本論ではハイパーゲームの形式による研究を取り上げる。

ハイパーゲームの枠組みでは、各エージェントは状況の内部モデルとしてゲーム形式を持つ。すなわち各エージェントはその状況において各自が採れる戦略の集合、他のエージェントが採りうると思う戦略の集合、そして各戦略に対して自己が得る利得（実数値）と他のエージェントが得られると考える利得とを利得関数として認識し内部モデル化する。注意すべきは、ハイパーゲームでは形式こそゲーム理論の用語を利用しているが、通常のゲーム理論とは異なり、各エージェントが見ているゲームが共通知識(common knowledge)として共有化されていない。内部モデルは各エージェントの状況認識を表している。

この形式において、他のエージェントの内部モデルとして記述されたゲームの要素である利得関数や戦略集合を遺伝的アルゴリズムを利用して各エージェントは上述した進化的方法の枠組みに基づいて学習する。

具体的にはたとえば複数のエージェントがネットワークを形成し学習を進めるネットワーク型の動的ハイパーゲームによる学習モデルが提唱されている。

エージェントは次のようなネットワーク型のハイパーゲームを構成する。各エージェントは自己の戦略集合と利得関数を持ち、他のエージェントの戦略集合と利得関数を認識する。エージェントはあるネットワークに所属する。ネットワークはいくつか存在する。ただし、現在までのシミュレーションでは二つのネットワークの場合が扱われている。

ネットワーク型ハイパーゲームは次のようなステップによる内部モデルの改善プロセスが実行される。

(1) プレイ。各エージェントは自己の所属するネットワーク以外のネットワークに所属するエージェントとゲームのプレイを行い、自己の採った戦略と相手が実際に採った戦略から自己の認識する利得関数に基づいて利得を得る。

(2) 結果の交換。各エージェントは自己の得た結果（自己の得た利得と相手が実際に採った戦略）を自己の所属するネットワークの他のエージェントと交換する。

(3) ゲームのルールの改善。各エージェントは交換した結果に基づいて自己の認識するゲームのルール（利得関数、戦略集合）を評価し、改善する。

(4) 新たなプレイ。改善されたゲームのルールに基づいて新たなプレイを行う。

これまでいくつかの設定において利得関数や戦略集合の学習実験が行われている。たとえば、採りうる戦略が  $a, b, c$  の三つで、それに対する利得が  $0, 1, 2$  のどれかであるような場合の実験では、エージェントは互いに相手の利得構造に関する内部モデルを持ち、それを学習する。これはこの特定の実験での設定であり、内部モデルの学習がこうでなければならないというのではない。またこの実験ではエージェントは利得最大を目指しており、最大利得である  $2$  が相手の利得関数のどの位置にあるかを学習を行う。

		Q					Q		
		a	b	c			a	b	c
P	a	2,*	1,*	1,*	P	a	*,2	*,0	*,1
	b	1,*	2,*	0,*		b	*,0	*,1	*,2
	c	0,*	0,*	2,*		c	*,1	*,2	*,0
エージェント P の認識する ゲーム状況					エージェント Q の認識する ゲーム状況				
(*印は P の認識する Q の利得)					(*印は Q の認識する P の利得)				

図 3 ハイパーゲーム状況の実験例

この実験から、内部モデルの学習に必要な情報のタイプに関する含意が得られた。情報は主に内部モデルの事後的評価に利用される。遺伝的アルゴリズムでは適応度による値の計算の部分である。適応度の計算ではいくつかの情報が利用されるが、内部モデルをうまく学習するには一種の場の情報が必要であった。場の情報は同一集団内のエージェント相互に情報交換することで得られる、相手の集団の各エージェントが採った手の一種の集計値である。これは今の状況に合わせて解釈された状況全体についての情報である。今回の実験例では、相手の集団の中でもっとも多く採られた手の、全体の手の数に対する割合（最頻値情報）と、相手の集団における採られた手の割合から導かれる可能な利得行列の形（分布情報）を利用している。最頻値情報は相手の状況に対する適応度を表わしている。すなわち相手の集団で正しい内部モデル持つエージェントが増加すれば最頻値の割合は 1 に近づく。分布情報はこちらの内部モデルと相手のエージェントの特性との整合性を表わしている。たとえば手 c を出すエージェントがいなかったとき、それに対する内部モデルとして手 c に最大利得 2 を配しているような個体は適応度が低くなる。

### 6. 3 システムモデル構築のためのシミュレーションの意義

上で述べた進化的方法の基本的ステップは、内部モデル学習のための一つの計算枠組みではあるが、本論で議論しているシステム理論的な説明モデルにはまだ距離がある。シミュレーションによりルール of 改善プロセスを実行することには、組織のシステムモデルを構築する立場から見たときいくつかの必然性と利点がある。組織学習は、エージェントの基本的行動ルールの相互作用から生み出される組織の創発的性質の一種である。すなわち、エージェントの持つルールに組織学習のためのプロセスをあらかじめ埋め込んでおけない。シミュレーションは、エージェントの行動ルールとエージェント間の相互作用からどのような改善プロセスが創発するかについての膨大な事象群を生成する。シミュレーションに

より生成された事象群は、一種の説明されるべき“データ”を提供してくれる。システム理論的に説明すべき対象は第一義的に実際の組織における実証的なデータとその性質ではあるが、少なくともシミュレーションによって生成された事象群とその生成メカニズムをも説明対象として含んでいなければならない。まずそれらを説明することで、そのシステムモデルの実際の組織事例への含意が得られていくであろう。すなわち学習プロセスの実現は、組織学習のシステムモデル構築にとっても重要である。

## 7. 組織の自己組織化と組織学習

最後に組織における自己組織化と組織学習との関係について簡単に触れておく。

システムの自己組織化はシステムの構造を自ら変動させて環境の複雑性（variety）に適応しようとする複雑システムの性質である。自己組織化の問題は複雑システムを考えるときには避けて通れないシステム理論の本来の問題の一つである。組織が複雑なシステムであり、環境の変化に適応するために自ら組織構造を変化させる必要性は広く認識されているといってもよい。しかしながらそのためのシステム理論的に十分な説明モデルが作られているとはいえない。

組織学習は環境に適応するための有効な組織行動の学習である。組織行動は組織構造に依存して生み出されるのであり、環境に適応する有効な組織学習のために組織構造の変動がともなうことは自然であり、必然ともいえる。組織学習の過程の解明は組織の自己組織化過程の解明につながる。

本論で提示した double-loop learning に対する進化的方法は、組織学習の具体的な過程の一つのモデルとなるものである。しかしながら、それがどのように組織構造の変動と関連しているかわかりにくいという見方もあるかもしれない。これに対する現段階での回答は次のようなものである。本論での枠組みでは組織は内部モデルを持ったエージェントのネットワークにより表現される。その構造は、各エージェントの内部モデルの構造とエージェント間のネットワークによる相互連関による構造により与えられる。double-loop learning による内部モデルの変化は内部モデルの構造を変化させることであり、内部モデルの共有化はエージェントネットワークの構造の変化に関連している。進化的方法はこれらの変化を各エージェントが自律的に、すなわち構造を変化させるためのコントロールセンターに依存することなく行う過程を明らかにしてくれるだろう。

## 8. おわりに

本論で述べたシステム理論の立場から組織論へアプローチは、説明のためのシステムモデルを開発することである。本論では組織論の問題として組織学習を取り上げ、そこでの double-loop learning を有効に行うための進化的方法の枠組みを提示した。しかし、未だそ

れは double-loop learning における内部モデルの改善過程を説明するシステムモデルとしての記述には程遠い。しかしながら、遺伝的アルゴリズムを用いたネットワーク型ハイパーゲームによる double-loop learning のシミュレーションの結果は、進化的方法が double-loop learning に対して十分有効であることを示している。また、エージェント間における知識共有化に関するシミュレーションからの知見は、組織学習ためのシステムモデルの開発への大きな試金石となるだろう。

エージェントの内部モデルの記述も本論で示したハイパーゲームに基づくものばかりでなく、他のモデルによる記述の可能性も探る必要がある。

システムモデルによる組織に関わる多様な事象の説明と理解は抽象度の高いものではあるが、それに対する適切な地位を確保する上でも、エージェントベースシミュレーションにより生成された事象群とデータを説明することが求められる。

## 参考文献

- Axelrod, R., *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton University Press, 1997.
- Barnard, C.I., "The Functions of the Executive," Harvard Univ. Press, 1938. (山本他訳「新訳経営者の役割」ダイヤモンド社, 1968.
- S.Beer, *Brain of the Firm*, 2nd ed., Chichester, 1981.
- Bowler, T.D., *General Systems Thinking, Its Scope and Applicability*, Elsevier North Holland, 1981. (T.ダウニング・パウラー「応用一般システム思考」中野文平訳, 紀伊国屋書店, 1983)
- Carley, K.M. and Gasser, L., "Computational Organization Theory," in *Multiagent Systems* (ed.: Weiss, G.), MIT Press, 1999.
- Checkland, P., *Systems Thinking, Systems Practice*, Wiley, 1981.
- Connant, R.C. and Ashby, W.R., "Every good regulator of a system must be a model of that system," *Int.J.Systems Sc.*, 1(2), 1970, pp.89-97.
- 出口 弘, "ネットワーク上の改善と進化の利得—その遺伝アルゴリズムによる基礎付—," *経営情報学会誌*, Vol.2, No.3, 1994.
- Dawid, H., *Adaptive Learning by Genetic Algorithms—Analytical Results and Applications to Economic Models—*, Springer, 1996, 1999(2nd ed.).
- Espejo, R., Schuhmann, W., Schwaninger, M. and Bilello, U., *Organizational Transformation and Learning—A Cybernetic Approach to Management—*, Wiley, 1996.
- Gaylord, R.J. and D'Andria, L.J., *Simulating Society—A Mathematica Toolkit for Modeling Socioeconomic Behavior—*, Springer, 1998.

- Hanappi,H., Evolutionary Economics, Avebury,1994.
- Holland,J.H., Adaptation in Natural and Artifitial Systems,MIT Press, 1992.
- 今井賢一, 金子郁容「ネットワーク組織論」, 岩波書店, 1988.
- 木嶋恭一編,マルチメディア社会システムの諸相,1997.
- Mesarovic,M.D.,Macko,D. and Takahara,Y.,Theory of Hierarchical, Multilevel Systems, NY:Academic Press,1975.
- Mesarovic,M.D. and Takahara,Y.,Abstract Systems Theory, Springer,1989.
- 森本三男,経営組織論,放送大学教育振興会,1987.
- 沼上幹,“われらが内なる実証主義バイアス” 組織科学,Vol.33,No.4,2000,pp.32-44.
- Prietula,M.J.,Carley,K.M. and Gasser,L.(eds.), Simulating Organizations, MIT Press, 1998.
- Putro,U.S., Kijima,K. and Takahashi,S., “ Adaptive Learning of Hypergame Situations Using a Gentic Algorithm,” IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans, Vol.30, No.5, 2000, pp.562-572.
- 塩沢由典,複雑さの帰結,NTT 出版,1997.
- 塩沢由典編,方法としての進化,進化経済学会,2000.
- 高橋真吾「システムの構造変動に対してゲーデルの不完全性定理の持つ意味」 第 17 回 計測自動制御学会システムシンポジウム講演論文集, 1991, pp.443-450.
- Takahashi,S.and Takahara,Y., Logical Approach to Systems Theory, Springer-Verlag,1995a.
- 高橋真吾, 中野文平, 荒瀬雅子「ハイパーゲーム的社会状況における認知の学習への遺伝的アルゴリズムの応用」 経営情報学会誌, Vol.4, No1,1995b,pp.43-56.
- 高橋真吾 「ハイパーゲーム型社会状況における学習：コミュニケーションによる情報生成」 進化経済学会論集第 2 集,1998,pp.314-322.
- 高橋真吾「組織の進化的学習のためのエージェント間インタラクションの基礎的考察」 経営情報学会秋季全国研究発表大会予稿集, 1999, pp.199-202.
- Takahashi,S.,Hinago,T.,Inohara,T. and Nakano,B., “ Evolutionary Approach to Three-person Hypergame Situation,” IEEE International Conference on Systems, Man and Cybernetics, 1999.
- Takahashi,S.,Evolutionary Learning in Agent-based Modeling,International Conference on AI,Simulation and Modeling,2000a(to appear)
- 高橋真吾「エージェントベースシステムのための数理的システム理論の課題」 社会・経済システム学会誌,2000b
- Takahara,Y., “A Mathematical Organization Theory,” Internal Report, Chiba Institute of Technology,1999.

- Ulrich,H. and Probst,G.J.B., "Self-Organization and Mangement of Social Systems," Springer,1984 (徳安 彰訳「自己組織化とマネジメント」東海大学出版会, 1992)
- Weibull,J.W., Evolutionary Game Theory, MIT Press,1996.
- Weiss,G.(ed.), Multiagent Systems-A Modern Approach to Distributed Artificial Intelligence, MIT Press, 1999.
- 山田誠二,適応エージェント,共立出版,1997
- B.Zeigler,H.Praehofer and T.G.Kim, Theory of Modeling and Simulation, Second Edition, 2000.

## Chapter 14

# Evolutionary Learning in Agent-based Modeling

S. Takahashi

*This paper develops a general model for evolutionary learning in agent-based modeling. The central concepts of the general model lie in internal model principle and mutual learning of agent's internal models in an evolutionary way. This paper particularly presents network-type dynamic hypergame as a model to describe an evolutionary learning process in multi-agent situation and a simulation method by genetic algorithm to perform a network-type dynamic hypergame. The experimental results given in this paper show some requisite conditions to progress the learning process effectively.*

### 14.1 Introduction

Agent-based modeling is basically a modeling method to reveal the nature of complex systems especially including complex social systems. The primal way of agent-based modeling is based on simulation of agents and their interactions involved in a complex system. One essential goal of agent-based modeling is to enrich our understanding of fundamental processes that may appear in a variety of applications for decision making.

Axelrod [1] insists that agent-based modeling is a third way of doing science, when it is contrasted with the two standard methods of induction and deduction. It is, at least, sure that agent-based modeling is becoming a powerful methodological tool of simulation to aid our intuition and decision making capability.

The aspects stated above give only a general notion of agent-based modeling. How each agent should be modeled depends on what we use as a specific formalism to describe the situation and agents.

“Evolutionary learning,” appearing in the title, mainly includes a comprehensive way of “evolutionary approach to learning process” by which we can understand the process that each agent involved in a complex situation learns in an evolutionary manner the situation and its relevant features. The term evolutionary approach indicates essential use of research



methods originated from Darwin's theory of evolution and its derived forms [9]. As an evolutionary learning method in simulation this paper will employ genetic algorithm developed by J.H.Holland [10].

The purposes of this paper are (1) to develop a general framework for evolutionary learning in agent-based modeling, (2) to introduce network-type dynamic hypergame to describe an evolutionary learning process based on game situation, and (3) to perform basic experiments of evolutionary learning by using genetic algorithm.

Recently, the concept and methodology of multiagent systems are playing one of central roles in many fields such as distributed artificial intelligence, organization theory, game theory or economics.

As an approach to distributed artificial intelligence, the area of learning in multiagent systems receives broad and steadily increasing attention [18]. Agents behave in intelligent or autonomous manners. We can distinguish various learning techniques provided in artificial intelligence and machine learning such as reinforcement learning or classifier systems based on genetic algorithms. They deal with learning models that basically maintain each agent's policy or the rule of actions [27].

The characteristics of agents modeled in this paper share mostly with the notion of agents formulated mainly in distributed artificial intelligence. Each agent in a complex system makes a decision autonomously according to the "rule" for agent's behavior. Furthermore this paper adds a more essential feature to agents: "internal model principle." The internal model principle requires each agent to have his own "internal model" of the situation. Each agent selects his action as a result of his decision based on both his internal model and the rule of behavior. Originally the internal model principle was introduced in the context of control theory [5]. In this paper we will extend and apply it to agent-based modeling, and clarify the relationship between internal model and the situation.

The internal model of an agent can be regarded as a model of "situation" in which the agent is involved or of "environment" with which the agent interacts. Hence evolutionary approach developed in this paper deals with the learning of the situation itself, while major learning techniques in distributed artificial intelligence focus on the learning of the rule of agent's actions.

On the other hand, economics also deals with adaptation process of bounded rationality, the rationality which assumes agents has insufficient ability to calculate optimal solutions. Several mathematical learning models have been provided such as least square learning [11] and Bayesian learning [3]. These models mostly assume a representative agent who represents a whole group of agents, improves his belief and react to his expectation based on some learning rule [6].

In this paper we formulate network-type dynamic hypergame to describe complex situations and internal models of agents to realize the general framework for evolutionary learning. Then we perform simulation of evolutionary learning process of the individual perceptions of the complex situation by applying genetic algorithm to a specific model of network-type dynamic hypergame.

The situation described in the network-type dynamic hypergame basically shows a game situation. In this paper such a situation is called a “hypergame situation.” It, however, has intrinsically different aspects from the “traditional” game situation [12,17] in the sense that each agent involved in the hypergame situation makes his decision based on the result from his “own” internal model of the situation.

The game situation is a framework for dealing with interactions of decision making of autonomous agents in social or economic system. The basic components of the framework of game are players, strategies, payoffs and preferences, in terms of which the situation is described. In a game situation each player has a set of strategies as options of actions, expects results when the strategies are selected, and gets some payoffs given by a payoff function assigning real numbers to strategies. According to the payoffs, the preference orderings are given. Each player plays a game by “rationally” selecting a strategy, which means he selects an optimal solution that maximized his payoffs.

A game can be formally defined by

$$\langle N, \{S_i, f_i\}_{i \in N} \rangle$$

where  $N$  is a set of players,  $S_i$  is a set of strategies of player  $i$  and  $f_i$  is a payoff function of player  $i$ , which assigns each strategy a real number. Nash solution is usually used as an optimal solution concept. The set of strategies  $(s_1, \dots, s_n)$  is said to be a Nash solution if for any  $i$  in  $N$  and any  $t_i$  in  $S_i$ ,

$$f_i(s_1, \dots, s_{i-1}, s_i, s_{i+1}, \dots, s_n) \geq f_i(s_1, \dots, s_{i-1}, t_i, s_{i+1}, \dots, s_n).$$

The Nash solution can be interpreted as the strategies from which every player has no motivation to move.

Hypergame [2,19] also describes the situation with the notions of players, strategies, payoffs and preference. In this sense, the hypergame situation can be said to be a kind of the game situation. However, in the hypergame situation no payoff functions are shared as common knowledge of the agents. This means that “the bounded rationality” is clearly and partly expressed in the hypergame situation.

On the other hand, the traditional game situation assumes a single description of the situation that is represented as a common payoff matrix, while in the hypergame situation each agent can have his own payoff matrix. From this point of view, evolutionary game theory [9,25] is based on “traditional” game theory and our evolutionary approach based on hypergame situation can be considered as another framework for learning of mutual perceptions.

In game theory [12,17] imperfect information game is also concerned with the situation where a player does not have exact knowledge about other player's payoff functions. However in game theory the imperfect information game is transformed into a Bayesian game, which is assumed as “equivalent” to the imperfect information game. Bayesian game is a “usual” game, in which any rules of game are provided as common knowledge among the players. In this sense, usual game theory does not primarily describe the hypergame situation.

This paper also presents a simulation model by genetic algorithm to perform network-type dynamic hypergame. The experiments in this paper

specifically perform the learning of the payoff matrices as agent's perception of the situation.

## 14.2 Internal Model Principle for Agent-based Modeling

Given a complex situation in which agents are involved, internal model is a representation of the situation created by an agent himself. Internal model is constructed “inside” the model describing agent’s behavior.

The notion of internal model was originally introduced in control theory. The main result about internal model is known as the internal model principle [5]: a control system should include a regulator isomorphic to the system to be controlled. Even when the environmental system for a control system should be controlled, the internal model principle is applicable and commutativity and faithfulness of dynamic properties of the control system should hold.

Complex social system comprises multiple agents who interact one another in making decisions. It is indispensable for each agent to “recognize” the characteristics of other agents about how they make their decisions and select their actions. This recognition of each agent can be represented as his internal model of the situation. The characteristics of agents in a particular decision making situation in agent-based modeling can be composed of, for example, alternatives the agents create, preferences among the alternatives, and so on. Hence an internal model in agent-based modeling should describe such characteristics as the rules of game in a game situation. Then an internal model works as a principle for each agent to make his decision. We should notice that an agent’s internal model is possibly different from other agent’s ones, whereas in usual game theory every agent has the same rule of game.

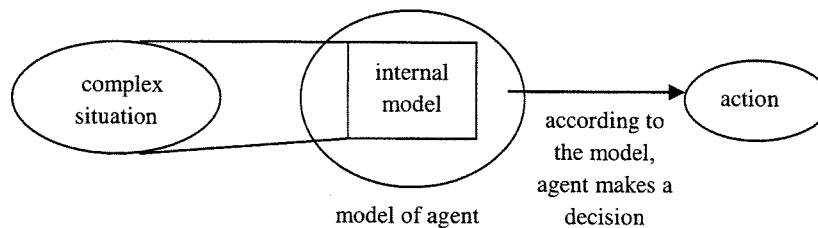


Figure 14.1: Internal Model

Each agent makes his decision according to his internal model. For example, when agent A prefers alternative y to x and believes that agent B prefers x to y, then agent A takes y based on his belief. If, however, agent B prefers y to x, then there must arise a conflict between agents A and B.

Internal model principle for agent-based modeling can be given as the phrase: an agent should construct his internal model isomorphic to the complex situation to make a “good” decision. This is our conjecture.

There are two aspects from which the homomorphism concept for validation in modeling methodology [8,22,23,28,29,30] is applicable to agent-based modeling. One is a homomorphism between an internal model of each agent and the situation. The other is a homomorphism between the model of agents and the reality. The usual concept of validation is concerned with the latter case. We should point out that the former case of “validation” of an internal model can be given by the “agent” based on only his “internal” criterion of decision principle, not by “us” as observers. An agent himself should examine the validity of his internal model, and build a valid internal model based on his decision and its response from the situation. In this sense learning process of internal models of agents would inevitably include the process of validation of each agent.

Based on the concept of internal model we need develop an effective way of actually constructing a “good” model isomorphic to the complex situation. One of the primary purposes of our research in agent-based modeling is to find a mechanism to evolve the internal models of agents. As a hopeful one this paper proposes an evolutionary way for such a mechanism.

### 14.3 Basic Steps and Method for Evolutionary Learning in Agent-based Modeling

This section clarifies the basic steps to be realized in simulation for evolutionary learning in agent-based modeling and provides belief sketch of genetic algorithm as a simulation tool.

#### 14.3.1 Basic Steps

The framework for evolutionary learning in agent-based modeling consists of the following four steps.

- (1) Each agent makes a decision and takes an action based on his decision independently.
- (2) Each agent exchanges the results as responses from the complex situation with other agents.
- (3) Each agent improves his “internal model” of the situation in some evolutionary manner.
- (4) Each agent makes a new decision and takes an action based on the improved internal model.

Since we assume hypergame situation, each of the above steps can be interpreted in term of game. In step (1) each agent selects a “strategy” so that the equilibrium can be achieved. In step (2) the only direct response from the situation is “opponents’ strategies” actually taken. Each agent also knows his own strategy taken and the payoff returned from the situation. We

should notice that an agent never knows a priori what payoffs his opponents actually obtained. In step (3) an agent's internal model can be represented by the components of a game; e.g. opponents' strategy sets and opponents' payoff functions. Then each agent improves his payoff matrix.

### 14.3.2 Genetic Algorithm

Genetic algorithm is a simulation tool originated from an evolutionary mechanism in biology [10]. Some essential concepts and words are borrowed from evolution theory: e.g. "chromosome," "gene," "selection," "fitness," "crossover," "mutation," and so on. The steps of simple genetic algorithm are

- (1) generation of initial population,
- (2) calculation of fitness value of individuals,
- (3) selection based on the fitness values,
- (4) crossover,
- (5) mutation, and
- (6) repeat from the step 2.

It is quite natural and important from several reasons that the genetic algorithm is applied directly to evolutionary learning in agent-based modeling.

First genetic algorithm can provide an effective computational way to perform the process of the improvement of internal models by distributive share and exchange of cognitive information of the complex situation. Furthermore genetic algorithm itself can be considered as a specific model of this process.

Second an actual learning process in genetic algorithm can be seen as a process of learning an internal model. Hence, in a sense, we can "observe" how the perception of the complex situation is improved, and an internal model adapts the situation.

## 14.4 Hypergame Model for Evolutionary Learning

A hypergame situation is basically described as a game situation in terms of the basic notions of "game," such as player, strategy and payoff. However the hypergame situation has essentially different features from the game situation.

In a hypergame situation each player behaves as an autonomous agent who makes a decision based on his own "internal model" that describes individually and independently the complex situation where he is involved.

This means that the rule of the game is not shared as the common knowledge among the players in a hypergame situation. Each component of the game is constructed individually based on each player's perception of the situation.

For example, let us describe a hypergame in international relations [11]. A and B are rulers of two nations, each desiring peace but suspicious of the

other. They both want mutual disarmament, but a search for security prompts an arms race as reacting to the threat of the other's weapons. They do not trust each other. We can model such a situation as a rather artificially simple one. We suppose that each player has the same strategies: arm or disarm. Figure 14.2 depicts the perceived strategies and their outcomes. A is supposed to love peace and have the following preference order:

Mutual disarmament > Arms lead for A (A arms, B disarms)  
 > Arms race  
 > Arms lead for B (A disarms, B arms).

This is not the situation B sees. He believes that A would most prefer an arms lead. Hence A's preferences "as perceived by B" are different from A's actual ones. Similarly A perceives B. Then the hypergame is described as in Figure 14.3.

		Nation B	
		Disarm	Arm
Nation A	Disarm	Mutual disarmament	Arms lead for B
	Arm	Arms lead for A	Arms race

Figure 14.2: Strategies and Outcomes

		B	
		Disarm	Arm
A	Disarm	4, 3	1, 4
	Arm	3, 1	2, 2

game perceived by A

		B	
		Disarm	Arm
A	Disarm	3, 4	1, 3
	Arm	4, 1	2, 2

game perceived by B

Figure 14.3: Hypergame: Example of "Mutual Suspicion"

If they both make their decisions and take their behavior according to their belief, then an arms race would be unfortunately selected, although the mutual armament should be realized as the most preferable alternative for both.

The hypergame above would help explain what was going on in the real situation. The problem we are concerned with is how both the nations could learn the situation or improve their belief to make a desirable decision. We should notice that the concept of hypergame situation does not properly include the learning process, but describe a complex situation to explain what was/is going on. To develop the learning process in hypergame situation we will define network-type dynamic hypergame after clarifying the problem of learning in hypergame situation.

A primitive description of hypergame situation is usually formalized as a simple hypergame. Here we call this model one-shot two-person hypergame since by essentially enhancing it we formulate an evolutionary model below as network-type dynamical hypergame.

**Definition 4.1 (One-shot Two-person Hypergame)** A one-shot two-person hypergame is defined by

$$(S_1, S_{12}, S_2, S_{21}, f_1, f_{12}, f_2, f_{21})$$

where

$S_i$  is the set of strategies of player  $i$ ,

$S_{ij}$  is the set of strategies of player  $i$  perceived by player  $j$ ,

$f_1$  is the payoff function of player 1 from  $S_1 * S_{21}$  to the set of real numbers,

$f_{12}$  is the payoff function of player 1 perceived by player 2

from  $S_{12} * S_2$  to the set of real numbers,

$f_2$  is the payoff function of player 2 from  $S_{12} * S_2$  to the set of real numbers,

$f_{21}$  is the payoff function of player 2 perceived by player 1

from  $S_1 * S_{21}$  to the set of real numbers.

In usual (traditional) game theory, non-cooperative two-person game is considered to be defined by  $S_i = S_{ij}$ ,  $f_i = f_{ij}$ . This means that in hypergame the recognition by player 1 is possibly different from the actual elements of the game of player 2. Then  $S_{ji}$  and  $f_{ji}$  represent the internal model of player  $i$ .

The problem concerning the learning process in agent-based modeling is to understand the process how each player can learn the rule of the game.

For example, in two-person hypergame given in Definition 4.1 the elements of the game, i.e. players, strategy sets, payoff functions, define the “situation” for players. In this situation we can say that player 1 completely “learns” the situation for him, if  $S_{21} = S_1$  and  $f_{21} = f_1$  are achieved.

We should notice that no player can know that the above conditions are actually achieved since player 1 can only perceive  $S_{21}$  and  $f_{21}$ . Hence such “correct” elements, i.e.  $S_{21}$  and  $f_{21}$ , cannot be used as the criteria of the player for the “learning” process in hypergame situation. The criteria of the player should be “internally” made from the actually perceived information and response from the situation.

There are two essential aspects to give an answer to the above learning problem in agent-based modeling.

One aspect is that the game should be repeatedly played, as the term “learning” implies “repeating.” Each game in one period of the repetition should be more improved than in the former period. In this sense this should not be called a “repeated game” that is performed with the same game in every period. We need to extend “one-shot” hypergame to more dynamic one.

The other essential aspect for learning in hypergame situation is “network.” The players in the situation should form a network with other players such as organization. As described above, a single player can have

only a restricted section of information of the complex situation. To produce effective information for learning the situation the players need to aggregate by “network” various sections of information gained from multifaceted aspects of them.

This shows a way of decision making by forming network. Autonomous decision makers form a network to exchange their sections of information and aggregate them. Then each decision maker improves his perception of the situation based on the result of the aggregation and makes the next step of decision. This type of decision making can be called “network-type decision making.”

Hence we extend one-shot hypergame to network-type dynamic one. First we define the network-type two-person hypergame.

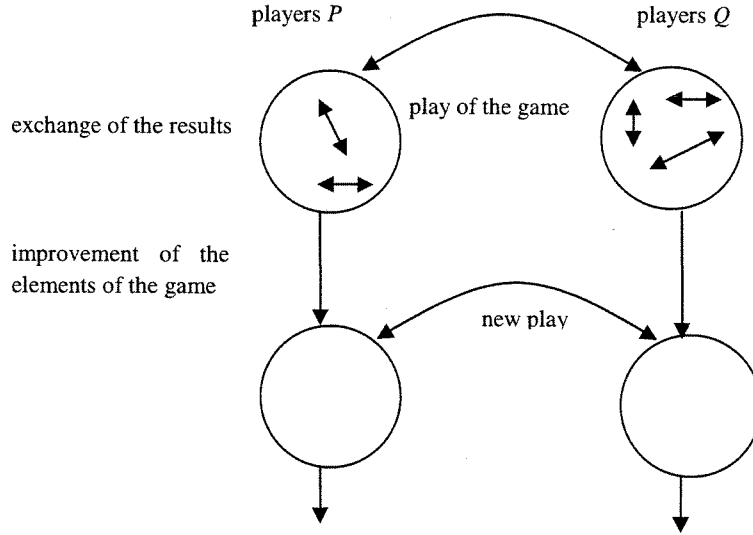


Figure 14.4: Network-type Dynamic Hypergame

**Definition 4.2 (Network-type Two-person Hypergame)** A network-type two-person hypergame is defined by

$$\langle \{P, Q\}, \{S_P, S_{Qp_i}, G_{p_i}, G_{Qp_i}\}_{p_i \in P}, \{S_Q, S_{Pq_j}, G_{q_j}, G_{Pq_j}\}_{q_j \in Q} \rangle$$

where

$P$  and  $Q$  are respectively the sets of players,  $P = \{p_1, \dots, p_m\}$  and  $Q = \{q_1, \dots, q_n\}$ ,

$S_P$  is the common strategy set of the players  $P$ ,

$S_{Qp_i}$  is the perceived strategy set of the players  $Q$  by player  $p_i$ ,

$G_{p_i}$  is  $p_i$ 's payoff function from  $S_P * S_{Qp_i}$  to the set of real numbers,

$G_{Qp_i}$  is the perceived payoff function of  $Q$  by  $p_i$  from  $S_P * S_{Qp_i}$  to the set of real numbers,  $S_Q, S_{Pq_j}, G_{q_j}, G_{Pq_j}$  are similarly given for  $Q$ .

The players in the same player set form a network to exchange the information.  $S_{Qp_i}$ ,  $G_{Qp_i}$ ,  $S_{Pq_j}$ , and  $G_{Pq_j}$  represent the internal models of player  $p_i$  or  $q_j$ .



The above network-type two-person hypergame expresses a one-shot game, in which only one play is performed. Based on this definition of network-type hypergame, we define the network-type dynamic hypergame to represent the learning process in hypergame situation.

**Definition 4.3 (Network-type Dynamic Hypergame)** A network-type dynamic hypergame is composed of a network-type two-person hypergame and the following steps.

1. **Play.** Each player in the player set plays with some player in the other player set and obtains the outcome and payoff.
2. **Exchange of the results.** Each player exchanges his results of the play with the players in the same player set.
3. **Improvement of the elements of the game.** Improvement of the perceived strategy set or the perceived payoff function. Each player changes and improves his perceived strategy set or payoff function so as that it is most plausible to interpret the exchanged results of the play.
4. **New play.** Then each player again plays with another player based on the newly perceived game.

A network-type dynamic game is evolutionarily repeated in the above way.

## 14.5 Simulation of Evolutionary Learning based on Network type Dynamic Hypergame

We demonstrate evolutionary learning process by some experiments based on the framework of network-type dynamic hypergame. First we put some assumptions to perform simulation.

### 14.5.1 Assumptions

This paper deals with the case where each player of  $P$  perceives a possibly different payoff function, but correctly perceives the strategy set of the opponent players. The player  $Q$  is supposed as the “nature” so that  $Q$  correctly perceives the situation.

We assume that  $S_P = S_Q = \{a, b, c\} = S_{Qp_i} = S_{Pq_j}$  and the payoffs of each player are either 0, 1 or 2. Hence each payoff function,  $G_{p_i}$ ,  $G_{q_j}$ , is a function from  $\{a, b, c\}^2$  to  $\{0, 1, 2\}$ , which might be different from the targets of the learning, i.e. it might be that  $G_{Qp_i}$  is not equal to  $G_{q_j}$ . These are parts of the common knowledge among the players.

In the current experiment, the payoff functions are given as follows.

		$S_{Qpi}$					$S_Q$		
		$a$	$b$	$c$			$a$	$b$	$c$
$S_P$	$a$	2,*	1,*	1,*	$S_{Pqj}$	$a$	*,2	*,0	*,1
	$b$	1,*	2,*	0,*		$b$	*,0	*,1	*,2
	$c$	0,*	0,*	2,*		$c$	*,1	*,2	*,0
		$G_{pi}$					$G_{qj}$		

Figure 14.5: Payoff Functions

The values of the asterisks (\*) in Figure 14.5 are given by  $G_{Qpi}$  and  $G_{Pqj}$  that are interpreted as “internal models” of the players.

Some specific interpretations of the above model are possible. For example, let us consider a marketing group that wants to know the behavior of consumers. Then the marketing group corresponds to the player set  $P$  and the consumers to  $Q$  as “nature.”

### 14.5.2 Steps of the Simulation

The basic steps of the current evolutionary simulation by genetic algorithm are as follows.

1. The simulator generates the player set,  $P$ , with the size of 30 populations. Each individual is expressed by

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9), x_i \in \{0,1,2\},$$

which are the elements of the perceived payoff functions,  $G_{Qpi}$  (see Figure 14.6).

		$S_{Qpi}$		
		$a$	$b$	$c$
$S_P$	$a$	2, $x_1$	1, $x_2$	1, $x_3$
	$b$	1, $x_4$	2, $x_5$	0, $x_6$
	$c$	0, $x_7$	0, $x_8$	2, $x_9$
		$G_{pi}$		

Figure 14.6: The game player P perceives

2. The strategy sets are specified. In the current simulation,  $S_P=S_Q=\{a,b,c\}$  is assumed to be correctly perceived by each player.
3. Each player of  $P$  plays with  $Q$ . Based on the individually perceived payoff function that is expressed as in the first step, one of the Nash solutions is performed in each play. To make Nash solution meaningful, we assume that each player “believes” that his internal model is correct and the opponent player also see the same payoff matrix. If the payoff functions have no Nash solution, then a strategy is randomly chosen. We should notice that any player cannot know which payoff his opponent has obtained. Each player can see only the strategy his opponent took, his own strategy and the payoff obtained.

4. The simulator calculates the fitness value of each player from the results of the play. The fitness, as stated later, is essential for the learning. It represents the characteristics of the exchange of information in the player set
5. Some players are selected as good parents using the proportional strategy of the fitness.
6. The one-point crossover is done. Then the perceived payoff functions are improved.
7. The next generation set of players are generated, and repeated from the 3rd step.

In the current experiments we did not use mutation, since it would not play any essential role in the results.

### 14.5.3 Fitness

In this section we describe the fitness used in the current experiments. From the result of a play in one period, each player can know directly the strategy what his opponent player took and his own payoff he got. The fitness is basically calculated from two values: opportunity loss and cost for changing the representation. The fitness  $f$  is defined as follows:

$$f = \frac{1}{1 + w * (\text{opportunity loss}) * (\text{cost for changing the representation})}$$

where  $w$  is a weight parameter.

The opportunity loss expresses the difference between the actually obtained payoff and the payoff that would have been obtained if the perceived payoff function were correct. In our case since each player plays a Nash solution, he could obtain the maximum payoff 2 if the perceived payoff function is correct. Then the opportunity loss can be calculated by 2 - (the actually obtained payoff).

The cost for changing the representation (payoff function) indicates a kind of penalty that should be paid in improving the payoff function from the results of the play. It is calculated basically by

$$ABS((\#existing \text{ Nash solutions}) - (\#expected \text{ Nash solutions})),$$

where  $ABS(*)$  stands for the absolute value of  $*$ ,  $(\#existing \text{ Nash solutions})$  is the number of the Nash solutions existing in the payoff function and  $(\#expected \text{ Nash solutions})$  is the number of the Nash solutions that are expected with some subjective probability. Here the probability is given according to the combinatorial structure of the payoff function: the probability that there is no Nash solution is  $0.296(=8/3^3)$ , the probability that there is one Nash solution is  $0.444(=4/9)$ , the probability that there are two Nash solutions is  $0.222(=2/9)$  and the probability that there are three Nash solutions is  $0.037(=1/3^3)$ .

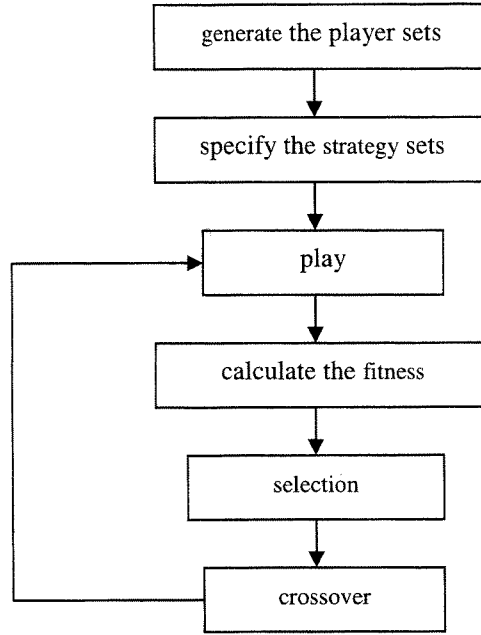


Figure 14.7: Steps of the Simulation

The following cases are separately calculated.

Case 1. The opportunity loss is zero and  $(\#existing\ Nash\ solutions) = (\#expected\ Nash\ solutions) = 2$ .

Then the cost for changing the representation takes 0 or 2 randomly.

Case 2. The opportunity loss is not zero, and

Case 2-1:  $(\#existing\ Nash\ solutions) = (\#expected\ Nash\ solutions) = 1$ , then the cost for changing the representation is given by 2, since we think that a Nash solution is not in the expected pair of strategies but in the unexpected one;

Case 2-2:  $(\#existing\ Nash\ solutions) = 1$  and  $(\#expected\ Nash\ solutions) = 2$ ,

then the cost for changing the representation takes 1 or 3 randomly;

Case 2-3:  $(\#existing\ Nash\ solutions) = 2$  and  $(\#expected\ Nash\ solutions) = 1$ ,

then the cost for changing the representation is given by 3;

Case 2-4:  $(\#existing\ Nash\ solutions) = (\#expected\ Nash\ solutions) = 2$ , then the cost for changing the representation takes 0 or 2 randomly.

#### 14.5.4 Selection and Crossover

In the 4th step of the simulation, we use the proportional strategy of the fitness to select two parents for the crossover. The probability  $P_i$  that an individual  $i$  is selected is given by

$$P_i = \frac{f_i}{\sum_{j=1}^n f_j},$$

where  $f_i$  is the fitness of  $i$ .

The crossover used in the 5th step is performed at one point chosen randomly. The children inherit the “genes” of one of the parents that have better payoff in the then play.

## 14.6 Results

The results of the experiments shown in this section are obtained by setting the parameters as follows. At first 30 sets of population are randomly generated. The probability of crossover is 0.6. The weight  $w$  varies from 0.1 to 10 by 0.1 step. These parameters were chosen based on the results of preparatory experiments done before the experiments. Under the same parameter conditions we repeated 3 times of simulation. Then in total we have 900 sets of players to be learned (30 sets of population\*10 weights\*3 repeats).

The results of learning fall into three categories as follows.

1. Learning the positions of Nash solutions in the payoff function (category 1).
2. Learning the positions of the maximum payoff 2 (category 2).
3. Learning the payoff function completely (category 3).

It seems impossible to achieve the learning of category 3 since in hypergame situation no player can obtain other information about the opponents than their taking strategies. Indeed from the results it is hard to distinguish the learning of category 3 from random search.

The learning of category 1 is weaker than that of category 2 in the sense that if the positions of payoff 2 are correctly perceived, then every Nash solution can be found. Our current experiment includes the case that there is no Nash solution in the payoff functions. Then category 2 is more suitable for the concept of learning in our case.

The following figures indicate the experimental results.

	Successful sets in number
Category 2	328/900
Category 3	47/900

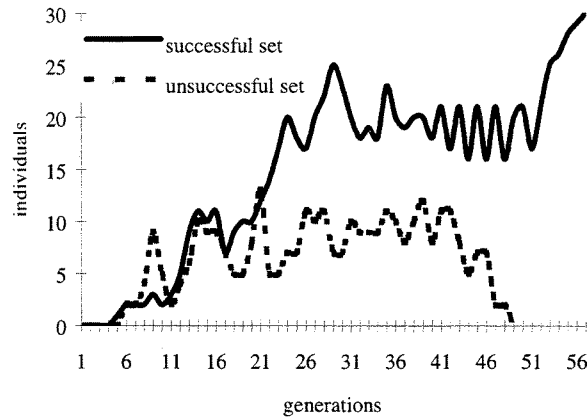


Figure 14.8: Typical Learning Process

When the weight is 0.1, the individuals succeeded most in category 2. Successful individuals in number decreased as the weight increased. In category 3 there was no remarkable difference among the weights. Though the successful learning processes depended on the patterns of the initial population sets generated in the first step, we did not see any specific one in the population sets. In particular, the experimental results did not depend on if the initial population set included the individuals having the same payoff with player  $Q$ . A typical learning process is shown in Figure 14.8.

The positions for the crossover used in this experiment were determined randomly. This caused fatal chromosomes, which had to be taken care separately. If the crossover is done so as not to cause fatal chromosomes, the result would be improved.

## 14.7 Conclusions

In this paper we developed a general framework for evolutionary learning in agent-based modeling. The framework consists of the four basic steps, in which the internal model and its improvement process are essential for the evolutionary learning. The internal model of an agent can be validated only by the agent's "internal" criterion of decision principle.

Then we formulated network-type dynamic hypergame to describe an evolutionary learning process based on game situation. A hypergame model is simple but very powerful to describe internal model of agent involved in the complex situation.

Based on network-type dynamic hypergame, we performed basic experiments of evolutionary learning by using genetic algorithm. The experiments given in this paper should be considered to provide a "basis" for further development of hypergame situation. There are a lot of directions to extend the current experiments based on the framework of evolutionary learning. A game situation can be recognized as a hypergame situation,

when more than one element of the rules of the game are not shared with players. The elements of the rules of the game include strategies, payoffs, players and outcomes. Then we can have many possibilities of problems: how agents learn about what strategies other agents have or how many agents are involved in the situation, and so on. So far some of extensions have been studied. They include the cases that two player sets learn “simultaneously” the other player’s payoff matrix [20], or that three player sets are involved in the situation [21], or that a player set learns the strategies of the nature under some restricted conditions on strategy [15,16].

Finally we should point out that organizational learning would be an important application field of evolutionary learning in agent-based modeling [4,7]. Organizational learning generally consists of two phases: learning by individuals and learning in organizational decision making by aggregation of individual capabilities. The concept of internal model can play an essential role in the both phases. In particular we expect that evolutionary approach would provide an effective method for the aggregation of internal models to extend the capability of organizational decision making.

## References

- [1] R.Axelrod, “The Complexity of Cooperation,” Princeton University Press, 1997.
- [2] P.G.Bennet, “Hypergames: Developing a Model of Conflict,” *Futures*, Vol.12, pp.489-507, 1980.
- [3] L.E.Blume and D.Easley, “Learning to be Rational,” *J. of Economic Theory*, 26, 340-351, 1982.
- [4] K.M.Carley and L.Gasser, “Computational Organization Theory,” in G.Weiss (Ed.), *Multiagent Systems – A Modern Approach to Distributed Artificial Intelligence*, MIT Press, 1999.
- [5] R.C.Connant and W.R.Ashby, “Every good regulator of a system must be a model of that system,” *Int.J.Systems Sc.*, 1(2), pp.89-97, 1970.
- [6] H.Dawid, “Adaptive Learning by Genetic Algorithms – Analytical Results and Applications to Economic Models,” 2nd Ed., Springer, 1999.
- [7] R.Espejo, W.Schuhmann, M.Schwaninger, and Bilello,U., “Organizational Transformation and Learning,” Wiley, 1996.
- [8] R.L.Flood and E.R.Carson, “Dealing with Complexity,” Plenum, 1988.
- [9] H.Hanappi, “Evolutionary Economics,” Avebury, 1994.
- [10] J.H.Holland, “Adaptation in Natural and Artificial Systems,” Ann Arbor: University of Michigan Press, 1975, (2nd ed) MIT Press, 1992.
- [11] A.Marcet and T.J.Sargent, “Convergence of Least Squares learning Mechanisms in Self Referential Linear Stochastic Models,” *J. of Economic Theory*, 48, 337-368, 1989.
- [12] A.Okada, “Game Theory,” Yuhikaku Press, 1996 (in Japanese).
- [13] C.Patterson, “Evolution,” British Museum (Natural History), 1978.
- [14] U.S.Putro,K.Kijima and S.Takahashi, “Simulation of Adaptation Process in Hypergame Situation by Genetic Algorithm,” SAMS, (to appear).

- [15] U.S.Putro,K.Kijima and S.Takahashi, "Simulation Approach to Learning Problem in Hypergame Situation by Genetic Algorithm," Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Vol.IV, pp.260-265, 1999.
- [16] U.S.Putro,K.Kijima and S.Takahashi, "Adaptive Learning of Hypergame Situations by Using Genetic Algorithm," IEEE Trans. of Systems, Man and Cybernetics (to appear).
- [17] E.Rasmusen, "Games and Information," Basil Blackwell, 1989.
- [18] S.Sen and Gerhard Weiss, "Learning in Multiagent Systems," in G.Weiss (Ed.), Multiagent Systems – A Modern Approach to Distributed Artificial Intelligence, MIT Press, 1999.
- [19] J.Rosenhead (ed.), "Rational Analysis for a Problematic World – Problem Structuring Methods for Complexity, Uncertainty and Conflict," Wiley, 1990.
- [20] S.Takahashi, "Evolutionary Approach to Hypergame Situation," IEEE Trans. of Systems, Man and Cybernetics (submitted).
- [21] S.Takahashi, "Evolutionary Approach to Three-person Hypergame Situation," Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Vol.IV, pp.254-259, 1999.
- [22] S.Takahashi, "General Morphism for Modeling Relations in Multimodeling," Transactions of the Society for Computer Simulation International, Vol.13, No.4, pp.169-178, 1997.
- [23] S.Takahashi and Y.Takahara, "Logical Approach to Systems Theory," Springer-Verlag, 1995.
- [24] S.Takahashi,B.Nakano and M.Arase, "Application of Genetic Algorithm to Analysis of Adaptation Processes of Individual Perceptions in Hypergame Situation," J.of Japan Association of Management Information, Vol.4, No.1, 1995 (in Japanese).
- [25] F.Vega-Redondo, "Evolution, Games, and Economic Behaviour," Oxford University Press, 1996.
- [26] J.W.Weibull, "Evolutionary Game Theory," MIT Press, 1996.
- [27] M.Wooldridge, "Intelligent Agents," in G.Weiss (Ed.), Multiagent Systems – A Modern Approach to Distributed Artificial Intelligence, MIT Press, 1999.
- [28] B.P.Zeigler, "Theory of Modelling and Simulation," Wiley, 1976.
- [29] B.P.Zeigler, "Multi-Faceted Modelling and Discrete Event Simulation," Academic Press, 1984.
- [30] B.P.Zeigler, "Object Oriented Simulation with Hierarchical, Modular Models: Intelligent Agents and Endomorphic Systems, Academic Press," 1990.



# 組織のダブルループ学習のための学習レベルを考慮した分配モデルによる基本的考察

## A Basic Model for Double-Loop Learning in Organization and Some Simulation Results by Genetic Algorithm

### Abstract

This paper presents a basic agent-based model for organizational learning. The model is constructed based on a distribution model of resources, and can distinguish the levels of learning in organization: single-loop and double-loop learning. In particular the double-loop learning can be specified as the parameters of other agents' characteristics that each agent recognizes. Then this paper uses genetic algorithm to simulate especially the double-loop learning by learning the agent's parameters of the model, and to clarify the properties of the successful fitness function and the model parameters.

### 1 はじめに

組織論におけるひとつの論点として、ミクロマクロループの形成に関する問題がある。現代の激変する環境に置かれた企業組織では、経済活動のすべてを市場に任せたり、あるいは反対に組織へと内部化したりするだけでは組織の成長だけでなく組織としての存続そのものが困難になっている。複雑な環境への適応に対して、“ネットワーク”という概念の有効性が以前から唱えられている。ネットワークの構築には、各構成員の自律的行動から組織全体への

リンクであるミクロマクロループ形成の過程が本質的であることも指摘されている。

組織学習の問題はミクロマクロループ形成の問題に関連した重要な論点のひとつである。組織学習は、組織の行動に有効性を与える組織の能力の向上と定義できる [9]。組織における高次のレベルでは環境の多様度と組織の多様度との調和を目指すことが一つの目標である。ではどのように環境または組織の多様度を有効に調節したらよいか。組織の学習は構成員個人の学習と本質的に連関している。個人の学習なくしては組織の学習はありえないし、組織の学習により行われる組織の行動が個人の行動や意思決定の制約条件ともなる。

この場合、ミクロループは single-loop learning と呼ばれ、構成員各自の学習であり、ミクロマクロループの形成は double-loop learning と呼ばれ、構成員の組織環境に対する解釈枠組み（メンタルモデル、内部モデル）の学習に相当する [9]。double-loop learning の本質は内部モデルの学習にある。そのとき double-loop learning を個人の内部モデル学習と組織による内部モデルの共有化とを考えることができる。Espejo et al.(1996) は前者を individual double-loop learning, 後者を organizational double-loop learning と呼んで区別している。

本論では組織における double-loop learning の進化的アプローチによる枠組みと、学習のレベル

を考慮した基本的な分配モデルによる double-loop learning の遺伝的アルゴリズムを用いたシミュレーション実験の結果を提示する。

## 2 エージェントベースによる組織学習の基本枠組み

本節で提示する組織学習のための基本的枠組みは以下ようになる。

1. エージェントはネットワークを形成している。
2. 各エージェントは状況の内部モデルを持つ。
3. 各エージェントは自己の内部モデルを参照して意思決定し行動する。
4. 自己の行動の結果と他のエージェントの行動は状況からのレスポンスとして観察され受け取る。
5. 観察し受け取ったレスポンスに基づき各エージェントは自己の内部モデルを修正する。

この枠組みの上で、組織学習の過程は次のように行われる。

### (1) individual single-loop learning

各エージェントが自己の内部モデルを参照して意思決定する過程である。エージェントが合理的意思決定をすると仮定するときは、individual single-loop learning は最適化の過程と一致する。

### (2) individual double-loop learning

各エージェントが状況からのレスポンスに基づいて自己の内部モデルを修正する過程である。

### (3) organizational single-loop learning

各エージェントの行動の集合による組織としての行動である。

### (4) organizational double-loop learning

各エージェントの内部モデルの共有化の過程である。

上記の組織学習の過程で最も問題となるのが二つの double-loop learning である。Individual double-loop learning では各エージェントが独立に自己の内部モデルを修正する。しかし、中立的な”正しい”モ

デルが存在してそれとの差を縮小するといった通常のコントロールの概念に沿った方法で修正をすることはできない。すなわち状況の”正しい”モデルを知らずに各エージェントは自己の内部モデルを修正しなければならない。さらに、修正された内部モデルが”正しい”かどうかの判断も基本的にはできない。organizational double-loop learning はさらに困難な過程のように見える。エージェントのネットワークにおいて内部モデルが共有されるとは、最も狭い意味では、各エージェントの持つ内部モデルが一致することである。各エージェントは自己の内部モデルを明示的に他のエージェントに示すことはしない。したがってエージェントのネットワークにおいてある一つのモデルが共有されたかどうかを各エージェントが個別に判断することは、内部モデルの比較という方法ではできない。内部モデルの共有化は各エージェントの内部モデルの修正の結果として得られる。すなわち有効な organizational double-loop learning の方法は、内部モデルの共有化を導くような individual double-loop learning、内部モデルの修正方法ということになる。

## 3 学習レベルを考慮した分配モデル

### 3.1 基本モデル

まず2人のエージェント、 $P_1, P_2$  からなる社会システムを考える。各エージェントは1つの意思決定変数を持ち、それは  $[0, 1]$  の閉区間内の実数値を取るとする。また、システムのアウトプットは2人の意思決定変数の値によって決定される。ここでは簡単のためにアウトプットは2人の意思決定変数の線形結合によって表されると仮定する。今はアウトプットの分配については問題となっておらず、両者間には分配の一定のルールがすでに存在している場合を考える。このルールは各エージェントの利得関数と通常呼ばれる。前節のエージェントベースの組織学習の基本枠組みおよび後述の GA によるシミュ

レーションからも  $P1, P2$  はおのおの組織メンバーによるネットワークを形成している集団である。

このとき各エージェントのモデルは次のように書ける。

$$P1: f_1 = a_1 u_1 + b_1 u_{21}$$

$$P2: f_2 = a_2 u_{12} + b_2 u_2$$

ここで、 $u_1$  はエージェント  $P1$  の意思決定変数、 $u_{21}$  はエージェント  $P1$  の予想するエージェント  $P2$  の意思決定変数、 $u_2$  はエージェント  $P2$  の意思決定変数、 $u_{12}$  はエージェント  $P2$  の予想するエージェント  $P1$  の意思決定変数、意思決定変数はいずれも  $[0, 1]$  区間内の実数値をとる。 $a_1, b_1, a_2, b_2$  は係数で実数値をとる。

システムのアウトプットが

$$z = a u_1 + b u_2$$

であるとき、各エージェントの利得関数の係数は分配のルールを表している。また、アウトプットの係数  $a, b$  はエージェントにとっての環境システムの構造を規定している。環境システムと分配のルールは全エージェントに最初から既知なのではない。それらは各エージェントがコストをかけて獲得する必要がある。それらがどうなっているかを「知る」ことは、組織の重要な活動であり、その判断は重要な経営戦略である。

### 3.2 学習の2つのレベル

上記の基本モデルにおける各エージェントの意思決定には2つのレベルが存在している。1つは、各利得関数を最大化する意思決定変数  $u_i (i = 1, 2)$  の決定と相手の意思決定変数  $u_{ij} (ij = 21, 12)$  の予測である。これはいわゆる single-loop learning に相当する。もう1つは、環境システム構造  $a, b$  と分配ルール  $a_i, b_i (i = 1, 2)$  の決定である。これは double-loop learning に相当する。環境システム構造と相手の持つモデルとを自己のモデル内にモデル化する内部モデルの学習となる。

学習のレベルを考慮したときの基本モデルは以下のようになる。

$$P1: f_1 = a_1 u_1 + b_1 u_{21}$$

$$P21: f_{21} = a_{21} u_{121} + b_{21} u_{21}$$

$$P12: f_{12} = a_{12} u_{12} + b_{12} u_{212}$$

$$P2: f_2 = a_2 u_{12} + b_2 u_2$$

ここで、 $a_{ij}$  はエージェント  $Pj$  が予測するエージェント  $Pi$  の持つ係数  $a_i$  の値である。他の係数についても同様である。 $u_{iji}$  は  $Pi$  が予測する「 $Pj$  の予測する意思決定変数  $u_i$ 」である。したがって、 $f_{21}$  はエージェント  $P1$  が持つエージェント  $P2$  に関する内部モデル、 $f_{12}$  は同様に  $P2$  の  $P1$  に関する内部モデルを表している。

### 3.3 モデルの特性

本モデルでは、通常のゲームは分配のルールが確定して各エージェントに共通に認識されているケースとして含まれる。たとえば  $a_{21} = a_2, a_{12} = a_1, b_{21} = b_2, b_{12} = b_1$  となっており、 $a_1 < 0, b_1 > 0, b_2 < 0, a_2 > 0$  の場合にはゲーム理論における囚人のジレンマと一致している。すなわち本モデルは通常のゲーム形式による状況記述に対し、分配ルールの学習レベルが導入されている。

状況を誤認識しているゲーム状況を扱うモデルとしてはハイパーゲーム [2] があるが、本モデルは相手の利得を誤認識したハイパーゲームを特別な場合として含む一般的な記述になっている。したがって、たとえば実際は囚人のジレンマの状況なのに、状況の認識を誤り囚人のジレンマとは認識していないようなハイパーゲームとしての解釈を本モデルによって行うこともできる。

さらに本モデルでは、 $f_i$  は  $u_i, u_{ji}$  に関する線形結合であるため、 $u_i, u_{ji}$  に関する単調関数となっている。たとえば、 $P1$  が  $a_1 < 0, b_1 > 0, b_{21} < 0, a_{21} > 0, |a_1| < |b_1|, |b_{21}| < |a_{21}|$  と状況を認識しているとき、 $f_1$  は  $u_1$  に関して減少し、 $u_{21}$  に関して増加する関数である。すなわち  $x \in [-1, 1]$  に対し、 $f_1(-1, u_{21}) > f_1(x, u_{21}) > f_1(1, u_{21}), f_{21}(u_1, 1) > f_{21}(u_1, x) > f_{21}(u_1, -1)$  となる。また  $P1$  は次のよ

うな選好順序が成り立つ。(これ以外の意思決定変数の値に関しては、 $f_1$  の単調性からこれらの中間に位置している。)

$$(-1, 1) > (1, 1) > (-1, -1) > (1, -1)$$

$f_{21}$  に関しては次のような選好順序が成り立つと  $P1$  は考えている。

$$(1, -1) > (1, 1) > (-1, -1) > (-1, 1)$$

これは  $P1$  が状況を囚人のジレンマと認識していることを意味している。しかし、 $P2$  の利得関数の  $a_2, b_2$  が  $P1$  の認識と異なっていれば囚人のジレンマは実際には起きていないことになる。

各エージェントの選好順序は上記のように環境状況を表す  $a, b$  の符号と大小関係により次のように定まる。表中の記号は、たとえば  $++>$  は環境状況が  $a > 0, b > 0, |a| > |b|$  であることを表す。

	選好順序
$++>$	$(1, 1) > (1, -1) > (-1, 1) > (-1, -1)$
$++<$	$(1, 1) > (-1, 1) > (1, -1) > (-1, -1)$
$+->$	$(1, -1) > (1, 1) > (-1, -1) > (-1, 1)$
$+-<$	$(1, -1) > (-1, -1) > (1, 1) > (-1, 1)$
$-+>$	$(-1, 1) > (-1, -1) > (1, 1) > (1, -1)$
$-+<$	$(-1, 1) > (1, 1) > (-1, -1) > (1, -1)$
$-->$	$(-1, -1) > (-1, 1) > (1, -1) > (1, 1)$
$--<$	$(-1, -1) > (1, -1) > (-1, 1) > (1, 1)$

表 3.1 環境状況の認識により決定される選好順序の組み合わせ

今回のモデルではエージェントは2人なので、表現できる状況は64通り ( $8 \times 8$ ) である。

各エージェントは各自の環境状況に対する内部モデルに基づいて最適な  $u_i$  を意思決定し (individual single-loop learning), その結果を各エージェント間で交換、共有して状況そのものの学習をして (organizational double-loop learning) 各自の意思決定を改善していくという過程をこのモデルでは基本的に描かれている。

## 4 GAによるシミュレーション

今回のモデルを使って、エージェントの double-loop learning である分配ルールの学習の遺伝的アルゴリズム (GA) によるシミュレーション実験を試みた。モデル上では2人のエージェント  $P1, P2$  は同時にしかも独立に学習する形式になっているが、今回の実験では  $P1$  のみが学習し、 $P2$  は正しく分配ルールを認知しているとする。また、 $P1, P2$  はエージェントのネットワークを形成し、その中で情報交換が行われる。情報交換は GA では適応度の計算に利用される情報として表現される。

### 4.1 手順

前節のモデルの特性で述べたように、環境状況は各利得関数の係数  $a_i, b_i (i = 1, 2)$  により決定される。環境状況は係数の大小関係により64通り存在する。今回のシミュレーションでは、各場合について  $P2$  は正しく認知し、 $P1$  のみに  $P2$  の係数を学習させる。各場合次のような手順で GA のアルゴリズムを適用した。

1. 集団のサイズは20とする。各個体は21ビットの2進文字列からなる遺伝子型を持つ染色体とする。閉区間  $[-1, 1]$  の実数値が表現型で、それへの変換は  $-1.0 + \frac{3}{2^{21}-1}x$  による。ただし  $x$  は2進表示の染色体を10進表示にしたものである。これにより、各染色体により小数点以下6桁までの実数値が表現される。
2.  $P1$  の各個体と  $P2$  とゲームのプレイをする。戦略はナッシュ均衡解とする。前節のモデルの特性により、利得は  $u$  が  $-1$  または  $1$  のとき最大となるので、ナッシュ均衡解も各エージェントの戦略が  $-1$  または  $1$  のときに与えられる。今回のモデルではナッシュ均衡解が常に存在する。
3. 各個体の適応度  $f$  を次により計算する。

$$f = \frac{1}{1 + w \times (\text{機会損失}) \times (\text{交換損失})}$$

		P2							
P1		++ >	++ <	+- >	+- <	-+ >	-+ <	-- >	-- <
	++ >	256/150	253/104	331/214	343/248	473/225	501/231	170/139	175/131
	++ <	236/121	258/127	331/241	333/241	504/259	496/247	163/128	171/108
	+- >	168/132	170/133	504/220	465/250	323/244	339/241	247/132	240/131
	+- <	183/134	163/130	490/264	504/254	337/253	336/248	235/118	231/104
	-+ >	235/121	245/125	415/262	399/251	503/235	465/274	204/134	198/120
	-+ <	233/141	252/130	420/250	396/224	500/262	507/284	206/125	204/138
	-- >	278/124	264/109	488/254	491/250	507/250	519/260	255/109	258/123
	-- <	170/135	150/122	521/238	494/262	333/228	340/271	228/133	250/117

表 4.1 P1 の学習結果

ただし,  $w$  は重み係数で, 機会損失は—(得られるはずであった利得)-(実際に得た利得)—により計算され, 交換損失は P2 の出した手をナッシュ均衡解とするために必要な行と列の入れ替え回数により計算される。

4. 適応度比例戦略を用いて次世代の親を選択する。
5. 与えられた交叉確率 (今回は 0.1) に従って一点交叉を行う。
6. 与えられた変異確率 (今回は 0.01) に従って突然変異を行う。
7. 新しい個体群を次世代として 2. から繰り返す。

環境状況 64 通りに対しておのおの 1000 回の試行をした。

## 4.2 結果

このシミュレーションでは適応度の役割が学習に果たす役割が本質的である。比較をするためにランダムに親を選択させた場合も実験した。表 4.1 は 64 通りの環境状況に対する P1 の試行 1000 回あたりの学習成功回数である。表中の記号は, たとえば ++ > は環境状況が  $a > 0, b > 0, |a| > |b|$  であることを表す。各状況で / の左側の数字が適応度比例戦略により親を選択した場合の成功回数で, 右側がランダムに親を選択した場合の成功回数である。

適応度比例戦略による選択とランダムな選択との間で成功回数に差があるかどうかを  $\chi^2$  適合度検定を行った。64 通りのうち 57 通りの環境状況について 1% 有意となり, 62 通りについて 5% 有意となった。有意とならなかったのは (++ >)-(-- >) と (-- <)-(++ <) の状況の 2 つのときである。

## 5 おわりに

本論では, まず組織における double-learning に対する進化的アプローチのための枠組みを提示した。このアプローチでは各エージェントの学習のレベルを考慮する必要がある。そのために本論では分配モデルを用いて各エージェントが意思決定する変数と, 分配ルールを定める利得関数の係数の決定とを分けることにより学習レベルを明示化した。

利得関数の係数の決定の学習を遺伝的アルゴリズムを用いて 2 つのエージェントネットワークが存在し, 一方は正しく状況を認知しており, もう一方が実際に環境を学習し double-loop learning する場合をシミュレーションした。

その結果今回導入した適応度関数によりほとんどの場合について有効な学習結果が得られた。しかし, ランダムな選択と有意な結果にならなかった 2 ケースについては適応度関数の改良によりさらなる研究が必要である。

## 参考文献

- [1] R.Axelrod, "The Complexity of Cooperation," Princeton University Press, 1997.
- [2] P.G.Bennet, "Hypergames: Developing a Model of Conflict," *Futures*, Vol.12, pp.489-507, 1980.
- [3] R.C.Connant and W.R.Ashby, "Every good regulator of a system must be a model of that system," *Int.J.Systems Sc.*, 1(2), pp.89-97, 1970.
- [4] U.S.Putro,K.Kijima and S.Takahashi, "Simulation of Adaptation Process in Hypergame Situation by Genetic Algorithm," *SAMS*, (to appear).
- [5] U.S.Putro,K.Kijima and S.Takahashi, "Simulation Approach to Learning Problem in Hypergame Situation by Genetic Algorithm," *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Vol.IV, pp.260-265, 1999.
- [6] G.Weiss (Ed.), *Multiagent Systems - A Modern Approach to Distributed Artificial Intelligence*, MIT Press, 1999.
- [7] S.Takahashi, "Evolutionary Approach to Three-person Hypergame Situation," *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Vol.IV, pp.254-259, 1999.
- [8] Dawid,H., *Adaptive Learning by Genetic Algorithms-Analytical Results and Applications to Economic Models-*, Springer, 1996, 1999(2nd ed.).
- [9] Espejo,R., Schuhmann,W., Schwaninger,M. and Bilello,U., *Organizational Transformation and Learning-A Cybernetic Approach to Management-*, Wiley, 1996.
- [10] Hanappi,H., *Evolutionary Economics*, Avebury,1994.
- [11] 木嶋恭一編, マルチメディア社会システムの諸相, 日科技連,1997.
- [12] Prietula,M.J.,Carley,K.M. and Gasser,L.(eds.), *Simulating Organizations*, MIT Press, 1998.
- [13] Putro,U.S., Kijima,K. and Takahashi,S., "Adaptive Learning of Hypergame Situations Using a Gentic Algorithm," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, Vol.30, No.5, 2000, pp.562-572.
- [14] 高橋真吾, 中野文平, 荒瀬雅子「ハイパーゲーム的社会状況における認知の学習への遺伝的アルゴリズムの応用」*経営情報学会誌*, Vol.4, No1,1995,pp.43-56.
- [15] 高橋真吾 「ハイパーゲーム型社会状況における学習：コミュニケーションによる情報生成」*進化経済学会論集第2集*,1998,pp.314-322.
- [16] 高橋真吾「組織の進化的学習のためのエージェント間インタラクションの基礎的考察」*経営情報学会秋季全国研究発表大会予稿集*, 1999, pp.199-202.
- [17] Takahashi,S.,*Evolutionary Learning in Agent-based Modeling*,in H.S.Sarjoughian and F.E.Cellier(eds.),*Discrete Event Modeling and Simulation: A Tapestry of Systems and AI-based Theories and Methodologies, A Tribute to the 60th Birthday of Bernard P.Zeigler*,Springer,2001.
- [18] 高橋真吾「エージェントベースシステムのための数理的システム理論の課題」*社会・経済システム学会誌*,2000.

# 進化的アプローチを構成する遺伝オペレータの組織システムへの解釈

Interpretations of Genetic Operators to Organizational Systems,  
the Operators which work as Essential Components in Evolutionary Approach

高橋 真吾

早稲田大学理工学部経営システム工学科

shingo@mn.waseda.ac.jp

**要旨** エージェントベースアプローチに基づくシミュレーションで期待される進化モデルの組織学習の過程への適用に際しては、その組織システムへの解釈の問題が組織システムの性質の解明に本質的である。

本論では遺伝オペレータを構成する4種類の進化モデルのコンポーネントを取り上げ、その組織での意味を探った。とくに、従来あまり議論されることがなかった“適応度”に関し、解釈上またシミュレーションの設計、実施上本質的役割を果たすことを議論した。また、“選択”は情報交換という解釈が重要である。一方で突然変異は意思決定者の意図しない過程であり、シミュレーション等の目的によっては省略可能である。

**Abstract** It is essential in revealing the properties of organizational systems to interpret substantially evolutionary models of organizational learning. The evolutionary approaches based on agent-based modeling can be considered to provide break-through methodology in systems research areas. This paper discusses on four genetic operators: population, fitness, selection and mutation, and points out that fitness plays an essential role in designing and interpreting a fitness function in an evolutionary simulation. An important interpretation of selection is exchange of information among agents within a decision making group. It is noticed that mutation is not indispensable for simulation experiments.

## 1 はじめに

組織等の複雑なシステムの性質、とくに環境適応の過程を解明する一つのアプローチとしてエージェントベースによるモデル化が最近注目をされている。たとえば、組織における学習問題を考察するために、組織と組織学習のモデルを作成し、その性質をコンピュータシミュレーションにより探る。

基本的に少なくとも二つの接近法がある。一つは、組織の当該の問題に関係があるパラメータをモデルに組み込み、“大きな”モデルを作ってシミュレートする。この長所は、モデルの意味が比較的取りやすく、結果の解釈が分かりやすいことである。しかし、モデルの動的性質はある程度モデル自体に最初から組み込まれていることがあり、次節で述べるように、シミュレーションによる問題点として指摘されることがある。もう一つは、当該の問題のうち、本質的なコンポーネントを成すサブシステムの構造を抽象的、一般的にまず構築し、その性質を調べる。この長所は、当該の問題の本質を比較的シンプルに把握できることである。しかし、欠点は対象への解釈が分かりにくい点である。

本論では、上記の第二の接近法としての進化的アプローチを組織の問題に適用した場合に生じる解釈の問題を扱う。とくに、進化モデルを構成

する本質的な遺伝オペレータについて、対象としての組織に対してどのような意味があるかを考察する。

## 2 モデルと解釈

モデルという場合には、モデル化された対象である何らかの原型が存在するはずである。しかし、モデル自体は原型とは“独立に”存在している。したがって原型を直接参照することなくモデルの内部のみで操作や推論ができる。また、モデルは原型の抽象化、一般化であり、その意味するところは“原型”へと解釈してやる必要がある。この“原型”は、モデル化された“元の”原型と同じである必要はない。

モデルの解釈は、大きく次の2つのカテゴリーに分けられる。一つは、モデル内部で推論された結論をモデル化のときに想定された“元の”原型の中で意味付けること。もう一つは、抽象化、一般化であるモデルを“元の”原型とは異なる“原型”において意味付けること。論理的にはもちろんこれらの複合的な解釈の方法がある。

シミュレーションモデルの場合、主に前者の解釈が行われる。すなわち、組織や社会等の考察の対象となる“原型”があり、それを元にモデル化し、シミュレーションによりモデルから一定の性質を導く。その性質を“元の”原型の中で解釈し、意味を探る。しかし、よく指摘されるシミュレー

ションモデルの欠点は、当初のモデルの中に、シミュレーションで導かれた性質をモデル化の時点で内在させている場合が少なくないことだ。これは知りたい性質、または論証すべき性質をあらかじめモデルの中に先取りした“論点先取”の誤りにつながる。

そこで、組織のシミュレーションモデルを構築する際に、原型である組織からの直接的な抽象化だけからでなく、他の原型からのモデルで、組織のダイナミクスを解明するのに必要な部分を持つものを組織において意味付けて組織に関する知見を得ることが重要である。すなわち、モデルの解釈の第二の方法である。本論で扱う解釈はこの場合である。原型は組織で、モデルは進化モデルである。

ある原型から直接別の原型へ解釈することは、ダイレクトアナロジーとも呼ばれ、理解しやすい反面、その意味付けが元の原型に引きずられやすい欠点がある。上記のモデル解釈の第二の方法はこのようなダイレクトアナロジーとは異なり、元の原型とは独立にモデルの解釈が行えるし、その必要がある。

### 3 社会・経済システムの進化モデルの利用類型

ダーウィンの進化理論に基礎を置く進化モデルが、社会・経済システムの適応過程を理解するためのモデルとして利用されてきている。これらにはモデルと原型との関係から見たとき、その利用のあり方にいくつかの類型がある。ここでは典型的な3つを考える。

#### 1. 進化理論を直接的に人間の社会・経済行動に適用

進化ゲーム理論を人間行動に適用し、人間の系統発生の初期段階として捉えるモデルや、ハイエクの社会的進化の理論がある。ハイエクの理論では、遺伝的進化の水準が“態度”の原始的形態の遺伝的固定にあるとし、情報伝達手段としての知識の進化や、行動に関し学習された規則の伝統としての文化的進化の水準について議論されている。しかし、淘汰の過程について十分な理解に至っていないとされる。

#### 2. 遺伝的進化と経済的進化との類推

進化理論を比喻として用いるときの問題は、淘汰の目的である。そして進化の実体は何かという問である。たとえば、生産物が手作りから大量生産へと“進化”した過程や、企業組織の成長、存続、現象市場からの退出への比喻である。これらへの問題点は淘汰機構の連続的作用を保証する遺伝プールが何かといったことや、遺伝型と表現型の区別が困難といったことが挙げられる。

もっと弱い形での類推は、たとえば Nelson and Winter が参考になる。遺伝型として、生産計画等のルーチンを取り、表現型は特定の諸決定とし、適応度を利潤の増加により表す。適応度が大きければそのルーチンの遺伝子の相対的頻度が増加する。しかし、この場合も連続的な淘汰機構の存在を示すまでに至っていない。

#### 3. 進化過程の抽象モデルの適用

進化理論の本質的要因を抽象化してモデル化したもので、現在有望なモデルの一つとして遺伝的アルゴリズムがある。特徴は、アルゴリズムが単純で、対象の表現次第で非常に広範な適用領域を持っていることである。遺伝的アルゴリズムが生物進化の過程を解明すると期待するのは困難であるが、他の領域にうまく適用することで、適用された対象の性質の理解を促進することは期待できる。

### 4 遺伝オペレータとその組織システムへの適用の持つ意味

#### 4. 1 対象とする遺伝オペレータ

本論で考察の対象とする進化的アプローチを構成する遺伝オペレータとして次のものを考える。

##### 1) 個体群、染色体、遺伝子

遺伝子には親の遺伝情報が直接乗っている。染色体は遺伝子から構成され、個体を表現し、個体群を形成する。

##### 2) 適応度

各個体は環境にどの程度適合しているかを適応度により数値化される。

##### 3) 選択

適応度に応じて次世代に子孫を残せる個体が



親として選択される。

#### 4) 交叉

選択された親は、子孫を残すためにペアリングされる。ペアとなった親の染色体が交叉により交換される。これにより、親の世代には存在しない遺伝情報を持つ子が生まれることがある。交叉は適応度の高い個体が選択された上で行われるので、子は親よりも適応度が高くなる可能性がある。交叉は本質的に新しい個体を生み出す過程となりえる。

#### 5) 突然変異

染色体上のどこかの遺伝子が何らかの要因によって他の遺伝子に変化することがある。これにより個体の形質が変化することもある。このような突然変異の起こる確率は一般には高くない。また、形質が変化するほどの突然変異が起きた場合には環境に適応できずに死に至る個体も出てくる一方で、環境にこれまでより適応する個体も出る可能性がある。ただし、交叉がかなり親の構造を引き継いだ系統的な変化であるのに対し、突然変異はランダムな様相が強い。

### 4. 2 解釈される進化モデル

さて、組織システムへ解釈される進化モデルを次のように考える。これはかなり一般化され抽象化されている。しかしそれだけに汎用性が高い。

複数の個体群が環境内にある。各個体群は個体から構成され、各個体は染色体により表現される。染色体は遺伝子から構成されている。各個体群の個体は適応度が計算され、適応度に応じて次世代の親の組が選択される。選択された親の組は染色体の特定のところで染色体が切れ、互いに遺伝子の交換を行い次世代の子とする。できた子は一定の確率によりその染色体上の特定の遺伝子に突然変異が生じる。新たにできた子の集団を次世代の個体群として、同様なプロセスが進行する。

### 4. 3 解釈の試み

#### 1) 個体群、染色体

シミュレーションモデルの立て方により、個体や個体群の意味するところが決まる。たとえば複数の個体群間でゲームのプレイを行い、結果を個体群内で情報交換するような場合、個体群が組織内のあるユニットを表現していると解釈するの

は自然である。複数の個体群全体の解釈には2通りある。一つは全体で一つの組織を表現している場合。たとえば組織内での分配モデルを基本にしたようなモデルが該当する。もう一つは、各個体群が別の組織を表現している場合。たとえば組織間での相互作用を前提としたモデルがある。いずれのモデルの場合も、現段階では組織全体を表現するほどの多様な相互作用を個体群間に持たせておらず、直接的な組織への解釈は単純な印象を受けることは否めない。しかしながら、本質的原理の追求のレベルとしては複雑あるいは大規模な全体組織のモデルよりかえって我々の理解を助けてくれる。

各染色体は、組織の各構成員を表すと解釈するのが自然ではあるが、むしろ、組織学習のような場合には、各個人の基本的行動類型あるいは基本的行動規則を表していると考えほうが有益である。

#### 2) 適応度

通常適応度は各個体の環境にどのくらい適応しているかを示すものとして、解釈そのものにはそれほど注意が向けられていない。しかし、これまでの著者らの研究では、この適応度の解釈が組織の学習過程の解明のために本質的な役割を果たしてきた。そこでは適応度の設計と解釈が研究の中心と言ってもよい。

組織における各意思決定者は基本的に自己の内部モデルを参照して意思決定する。これは他者に持つ選好や利得など、本来他者の持つ情報は直接所与ではないことを意味する。意思決定者は限定合理的であり、利用可能な情報が限られている。さらに、自己の内部モデルが環境にどの程度適応しているかの判断は、各意思決定者が限られた情報から行わねばならない。最適化問題における所与の目的関数が存在しない状況である。

適応度は、各意思決定者が限定的に認知している情報に基づいて、現在の自己の内部モデルの評価を表現している。分析者の立場からは、各個体の情報がすべて見えているが、これらを各個体の適応度に反映させるような適応度の設定はしてはならない。あくまで、適応度を構成する情報は各個体が認知できる情報に限定しなければならない。そうすることで、学習に必要な情報の性質が浮き彫りになる。

一方で、このように情報の解釈に重点を置いた適応度の設定では、常に学習に成功するといえる実験を再現させるのに骨を折ることになる。状況を少し複雑にすると、実験そのものが学習に成功しているかどうかを統計的に検定する必要が出てくる場合もある。

### 3) 選択

選択は、大きく二つの解釈があり得る。一つは、自己の属する集団内の他の意思決定者の行動様式を“まねる”効果を表している。つまり、よい適応度を持つ個体の遺伝子が他の個体に伝わる。もう一つは、意思決定者間のコミュニケーションと情報交換の効果を表している。各意思決定者の獲得する情報は環境構造の部分的な側面を表しているに過ぎない。それらを集団内で情報交換し、共有していく過程が“選択”である。

### 4) 突然変異

これはランダムに生起するところが本質である。すなわち、意思決定者の意図しない過程である。しかし効果は各意思決定者の内部モデルに現れ、意思決定に影響する。もっとも自然で直接的な解釈は、各意思決定者の認知の表現時での“意図しない”誤りと、行動の際の“意図しない”誤りである。これは各自の内部モデルが環境構造を“正しく”認知していないという意味での“系統的”誤認知とは本質的に異なる。学習が決定的に改善されるのでなければ、突然変異を実験過程に原理的に入れる必然性はない。

## 5 おわりに

エージェントベースアプローチに基づくシミュレーションで期待される進化モデルの組織学習の過程への適用に際しては、その組織システムへの解釈の問題が組織システムの性質の解明に本質的である。

本論では遺伝オペレータを構成する4種類の進化モデルのコンポーネントを取り上げ、その組織での意味を探った。とくに、従来あまり議論されることがなかった“適応度”に関し、解釈上またシミュレーションの設計、実施上本質的役割を果たすことを議論した。また、“選択”は情報交換という解釈が重要である。一方で突然変異は意思決定者の意図しない過程であり、シミュレーショ

ン等の目的によっては省略可能である。

## 参考文献

1. Carley, K.M. & Gasser, L., “Computational Organization Theory,” in Multiagent Systems (ed.: Weiss, G.), MIT Press, 1999.
2. Takahashi, S., Evolutionary Learning in Agent-based Modeling, in H.S. Sarjoughian and F.E. Cellier (eds.), Discrete Event Modeling and Simulation: A Tapestry of Systems and AI-based Theories and Methodologies, A Tribute to the 60th Birthday of Bernard P. Zeigler, Springer, 2001.
3. 木嶋恭一編, マルチメディア社会システムの諸相, 日科技連, 1997.
4. 高橋真吾, 中野文平, 荒瀬雅子, “ハイパーゲーム的社会状況における認知の学習への遺伝的アルゴリズムの応用,” 経営情報学会誌, Vol.4, No.1, 1995, pp.43-56.
5. Espejo, R., Schuhmann, W., Schwaninger, M. and Bilello, U., “Organizational Transformation and Learning,” Wiley, 1996.
6. Putro, U.S., Kijima, K. and Takahashi, S., “Adaptive Learning of Hypergame Situations Using a Genetic Algorithm,” IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, Vol.30, No.5, 2000, pp.562-572.
7. 高橋真吾, “組織の進化的学習のためのエージェント間インタラクションの基礎的考察,” 経営情報学会秋季全国研究発表大会予稿集, 1999, pp.199-202.
8. 出口 弘, “ネットワーク上の改善と進化の利得—その遺伝アルゴリズムによる基礎付け—,” 経営情報学会誌, Vol.2, No.3, 1994.
9. Axelrod, R., “The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration,” Princeton University Press, 1997.
10. Hanappi, H., “Evolutionary Economics,” Avebury, 1994.
11. 高橋真吾, “エージェントベースシステムのための数理的システム理論の課題,” 社会・経済システム学会誌, 2000.
12. 高橋真吾, “組織のダブルループ学習のための学習レベルを考慮した分配モデルによる基本的考察,” 計測自動制御学会, 自律分散システムシンポジウム, 2001.
13. Takahashi, S., Hinago, T., Inohara, T. & Nakano, B., Evolutionary Approach to Three-person Hypergame Situation, IEEE International Conference on Systems, Man and Cybernetics, 1999.
14. Takahashi, S. & Takahara, Y., “Logical Approach to Systems Theory,” Springer, 1995.

# 相互認識を考慮した一般モデルによる組織学習レベルの考察

## A General Model of Mutual Recognitions of Agents that Expresses the Levels in Organizational Learning

高橋真吾

Shingo Takahashi

早稲田大学 理工学部 経営システム工学科

Department of Industrial and Management Systems Engineering

Waseda University

**ABSTRACT** This paper presents a general model describing the elementary four organizational learning loops, which were originally introduced by Argyris and Schon. The formulated model in this paper can distinguish the levels of organizational learning in a general way. The central problems in organizational learning lie in the processes of revising agents' internal models and sharing them in an organization. This paper provides essential steps of a revising and sharing processes of agent's internal models in evolutionary approaches by using genetic algorithms.

### 1 はじめに

組織学習は、1960年代から主に経営組織論から登場し、80年代後半から現在に至るまで盛んに研究されてきている。本稿の組織学習に対する基本的スタンスは、いわゆる Argyris 系といわれる一連の研究の流れ[1]に沿うもので、組織の構成員を個々の意思決定主体として扱い、それらの相互作用から組織全体に関わる価値および解釈枠組みの変化に焦点を当てている。

従来の研究では、個人の学習がどのように組織全体の学習へと繋がっているのかの解明が必ずしも十分とは言えない。そのためのひとつのアプローチとして、両者を関連づけるプロセスのモデルを操作的なレベルにおいてエージェントベースの観点から構築することがある[3,4]。

そのモデルでは、個人の学習と組織全体の学習とが明示的に同定できるレベルとして記述されている必要がある。また、このモデルは組織学習の基本的メカニズムを記述するものであり、組織の包括的機能の観点から記述した“組織モデル”の中では一種の隠れた機能モジュールの役割を果たすことになる。

本稿では、組織全体の学習を、個人の状況に対する内部モデルの共有化のプロセスによる組織への解釈枠組みの変化ととらえる。その共有化のプロセスを記述

するための基礎的なモデルを経済学における分配モデルに基づいて提示する。このモデルでは、個人の学習のレベルと、組織全体の解釈枠組みの変化である内部モデルの変化と共有化のプロセスとしての学習レベルとを明示的に表現できることを考察する。

上述のような本稿の主題の目標は、後述するように、組織学習のとくに double-loop learning と呼ばれるメカニズムを、数理モデルのレベルで理解することである。そのためには解析的およびシミュレーション的接近の両方が必要とされる。本稿では、これまでの結果と今後の方向についても触れる。

### 2 組織学習における4つの learning loops

本稿が組織学習の議論の出発点としている枠組みは、Argyris らによるもので、個人の学習の観点から組織への学習へと連関させた視点を有している。そこでは、個人レベルと組織レベルが区別され、さらに各々 single-loop と double-loop という学習レベルが区別される。合計4種類の learning loop の概念が区別される。本稿では、個人の持つ内部モデルの学習の観点から learning loop の特徴を述べる。この観点は主に Espejo の組織サイバネティクスによる組織学習の考え方[16]にその基礎がある。

Argyris らの single と double の学習レベルの区別[9]は、以下のように Ashby のサイバネティックな考え方[6]に基づいている。学習レベルに関する本質的なことから含むので、少し長いが引用しよう。

We borrow the distinction between single- and double-loop learning from W.Ross Ashby's design for a brain[6]. Ashby formulates his distinction in terms of (a) the adaptive behavior of a stable system, "the region of stability being the region of the phase space in which all the essential variables lie within their normal limits," and (b) a change in the value of an effective parameter, which changes the field within which the system seeks to maintain its stability. One of Ashby's examples is the behavior of a heating or cooling system governed by a thermostat. In an analogy to single-loop learning, the system changes the values of certain variables (for example, the opening or closing of an air valve) in order to keep temperature within the limits of a setting. Double-loop learning is analogous to the process by which a change in the setting induces the system to maintain temperature within the range specified by a new setting. [9,pp.21]

## 2.1 individual single-loop learning

これは個人の学習レベルのひとつである。このレベルでは、意思決定者はその内部モデルを参照して意思決定を行う。内部モデルは修正を受けない。内部モデルの中で操作される意思決定変数のみを変更して環境に適応する。

この学習レベルは、上述の引用からもわかるように、サイバネティクスの考え方による、設定された目標値とのずれを修正する過程で、負のフィードバックとして捉えられる。サーモスタットのような単純な制御過程や、合理的意思決定過程における最適化等も含まれる。過去の事象を“利用する”し、未来の事象を“探求する”という意味では“学習”とも言えるが、基本的には一定の基準のもとでの意思決定プロセスであり、この学習レベルを“learning”と呼ぶのが適切かどうか

の問題は残る。

## 2.2 individual double-loop learning

これは個人の学習レベルのうち、内部モデルにおける環境パラメータに相当する変数を変更して環境に適応する。これにより内部モデル自身が修正を受ける。修正方法をどのようにするかが中心の問題となる。

## 2.3 organizational single-loop learning

組織全体の single-loop learning は個人の single-loop learning の合成として実現される。組織全体のパフォーマンス向上のための負のフィードバックのための修正過程は、各意思決定者の内部モデル内の意思決定変数の修正過程の統合として行われる。

## 2.4 organizational double-loop learning

この学習レベルで、組織全体の状況に対する解釈枠組みの変更と解釈できる過程により、組織のパフォーマンスが向上する。内部モデルは個々の意思決定主体が持つものであり、組織の持つ内部モデルというものは個人とは独立には存在しない。

このレベルの学習においても、個人の double-loop learning の統合として、各意思決定主体の持つ内部モデルが組織内で共有化されることで実現される。これはサイバネティックな観点からは一種の second-order cybernetics とも考えられるだろう[16]。

各意思決定主体は、organizational double-loop learning の結果共有化された内部モデルについて、共有化されたという状態を直接認知することはできない、と考えるのが自然である。

## 3 相互認識を表現する一般分配モデル

本節では、前節で述べた4つの learning loop を表現し、とくに double-loop learning のメカニズムを探るために、一般的なモデルの定式化を試みる。

### 3.1 一般分配モデルの定式化

エージェント集団は  $n$  人からなる。各エージェントは意思決定変数  $u_i$  を持つ。各  $u_i$  について係数パラメータ  $a^i$  があり、システム全体のアウトプットは

$$z = \sum_i a^i u_i, \quad (1)$$

で表される。各エージェントは自分の意思決定変数  $u_i$  を操作することで意思決定を行い、それによりシステム全体のパフォーマンスが決定する。各エージェントは全体のパフォーマンスから分配を受ける。分配は各エージェントに対する分配係数に基づいて行われる。また、各エージェントが受け取る分配は、他のエージェントの意思決定変数の値と自分の意思決定変数の値に各々分配係数を掛けて和を取ったもので与えられる。すなわち、

$$f_i = \sum_k a_k^i u_{ki}, \quad (2)$$

変数  $u_{ki}$  は、エージェント  $i$  が予測するエージェント  $k$  の意思決定変数を表している。ただし、 $u_{ii} \triangleq u_i$  と表すことにする。

このモデルではまだエージェントの内部モデルが表現されていない。すなわち各エージェントが認識する他のエージェントの分配関数である。  $i$  の認識する  $j$  の分配関数は次のように表現できる。

$$f_{ji} = \sum_k a_k^{ji} u_{kji}, \quad (3)$$

ここで、 $a_k^{ji}$ 、 $u_{kji}$  は各々  $i$  の認識する  $j$  の分配係数と  $i$  の認識する  $j$  の  $k$  について予測した意思決定変数を表す。ただし、 $u_{jji} \triangleq u_{ji}$  とする。

### 3.2 階層的エージェントの表現への拡張

上記の一般分配モデルでは、エージェント集団は  $n$  人から成り立っているとした。しかし、この表現のみでは、組織をフラットな構造としてしか捉えていない。また、3.5 節のシミュレーションによる方向のところでも述べるように、組織学習のメカニズムに接近するためにも、階層化した組織の表現に拡張する必要がある。

ここでは、最小のモジュールとなることを考慮して、2 階層としてモデル化する。すなわち、エージェントグループの集合  $N = \{1, 2, \dots, n\}$  を考える。各  $i \in N$  はエージェント集団  $i = \{i_1, \dots, i_p\}$  を表す。各  $i_p$  は式(2),(3)の形式の分配関数を持っている。すなわち、

$$f_{i_p} = \sum_k a_k^{i_p} u_{ki_p}, \quad (2)'$$

$$f_{ji_p} = \sum_k a_k^{ji_p} u_{kji_p}, \quad (3)'$$

$i_p$  の認識は、同一のエージェントグループに属するエージェントに対しては同一であるとしている。ただし、同じエージェントグループに属している 2 人のエージェントの同一のグループに対する認識が異なっていることを許している。

この階層化された一般分配モデルでは次の 2 通りの解釈が可能である。

- 1) 各エージェントグループは同一組織内の“部門”のようなサブ組織を表わし、各エージェントはどこかの部門に排他的に属している。各エージェントは他の部門に属するエージェントの認識を部門に対するものとして持っている。
- 2) 各エージェントグループは異なる組織を表わし、各グループのエージェントは各々別の組織の構成員である。このときは、システム全体のパフォーマンスを表す(1)式は必ずしも必要ない。

どちらが適しているかは、問題状況の設定に依存する。注意するのは、この拡張されたモデルにおいては、まだ組織学習のメカニズムは記述されていないことである。それはまた新たなアプローチを必要とする。3.5 節で述べる。

### 3.3 モデルにおける学習レベルの表現

本稿の目的のひとつである、学習レベルの表現が定式化された一般モデルにおいてどのように記述されるか考察しよう。

エージェントが  $i_p$  毎回の意思決定において扱う直接的な意思決定変数は  $u_{i_p}$  である。また予測するのは他のエージェントの意思決定変数  $u_{ji_p}$  である。自己の分配（パフォーマンス）を目標値（通常最大化）にするために、予測値を見極め、自己の意思決定変数を決定

する。このプロセスは 2 節で述べた individual single-loop learning に相当する。

これらが合わさってエージェントグループおよび全体組織の意思決定が行われている。このプロセスは organizational single-loop learning になる。

分配係数は各エージェントによって認識されるが、これは操作的な意思決定変数ではなく、意思決定を制約している環境状態の認識を表している。すなわち、

分配係数  $a_j^{i_p}, a_j^{j_p}$  はエージェントにとっての環境であ

り、エージェントが操作するものではない。エージェントが操作するのは、環境の“認識”記述である。この認識を変更することは、自己の意思決定変数  $u_{i_p}$  の適

用領域を大きく変化させることになる。言い換えれば、適用領域の変化によって環境に適応するプロセスである。つまり、これは individual double-loop learning に相当する。

したがって、エージェントグループにおいて分配係数が共有化される過程が organizational double-loop learning を表す。

以上をまとめると、各 learning loop とモデルにおける変数との関係は次のようになる。

#### 1) individual single-loop learning

$u_{j_p}$  を予測し、自己の受け取る分配が目標値となるように自己の  $u_{i_p}$  を決定する。

#### 2) individual double-loop learning

前回の自己の意思決定変数  $u_{i_p}$  と、他のエージェントが実際に選択した意思決定変数、および前回自己が得た分配から、あるメカニズムにしたがって、自己の環境認識である  $a_j^{i_p}, a_j^{j_p}$  を修正する。

#### 3) organizational single-loop learning

目標とするパフォーマンスの実現のために、各エージェントに対してサブ目標が伝えられ、

individual single-loop learning に基づく各エージェントの意思決定の結果、組織全体のパフォーマンスが得られる。

#### 4) organizational double-loop learning

individual double-loop learning の結果、エージェントの他のエージェントの環境認識  $a_j^{j_p}$  がエージェントグループ内で共有化される過程である。

### 3.4 モデルの数学的特性

モデルの数学的特性として、ここで指摘しておきたいことは、本稿で定式化したモデルは、いわゆるハイパーゲームによる表現を特殊な場合として含んでいることである。

実際、エージェントの意思決定変数の取りうる範囲と分配関数の特性を限定すれば、囚人のジレンマ状況を誤認識しているようなハイパーゲームを表現することができる[23]。

### 3.5 シミュレーションの方向：進化的アプローチ

本稿で定式化した一般分配モデルで、組織学習のレベルが表現できることがわかるが、モデルでは表現できていない重要な部分が2つある。1つは individual double-loop learning における環境認識（エージェントの内部モデル）の修正メカニズムで、もう1つは、organizational double-loop learning における環境認識（内部モデル）の共有化の過程である。

このメカニズムや過程に含まれる本質的困難は、次のステップでの変化先、さらには変化の領域すらあらかじめ特定化できないことである。したがって、数理的定式化でよく行われるような、メカニズムやプロセスを一種の関数によって表現することは、先験的に変化領域を設定することになり、今の場合問題の解決にはならない[26]。

現段階では修正メカニズムや内部モデル共有化の数理モデルによる記述はできていないが、それらの解明に有力な方法として、進化的アプローチによるシミュレーションを利用した研究がこれまで進められている。具体的には遺伝的アルゴリズムを利用するが、本稿の

ような適応システムに対してはその有効性が指摘されてきている[7,15]し、筆者らの研究でも確かめられている[11,12,20,21,22,24]。

ここでは遺伝的アルゴリズムを使ったときの本質的ステップを述べる。

- 1) エージェントはネットワークを形成している。
- 2) 各エージェントは状況の内部モデルを持つ。
- 3) 各エージェントの内部モデルを染色体として表現する。
- 4) 各内部モデルに関する適応度を計算する適応度関数を設定する。
- 5) 自己の行動の結果と他のエージェントの行動は状況からのレスポンスとして観察され受け取る。
- 6) 観察し受け取ったレスポンスのみに基づき各エージェントの内部モデルを事後評価する。
- 7) 事後評価に基づいて、選択、交叉、突然変異といった遺伝的オペレータを作用させ、内部モデルを修正する。

上記においては、適応度関数の設定が最も重要な鍵である。ここでは、各エージェントが利用できる情報のみを利用して適応度を計算する必要がある。なぜなら、ステップ7)での遺伝的オペレータは基本的には各エージェントにより自己の内部モデルに対して作用させられるものだからである。

本稿の階層化された一般モデルでは、エージェントグループがGAにおけるエージェントのポピュレーションに相当する。このとき、上述のステップにより、複数のエージェントグループが同時に、独立に適応する。この同時適応については、筆者らによるハイパーゲームの枠組みのもとでの研究が存在する。しかし、本稿のモデルでは、一つのエージェントグループが適応する場合のみについて結果[5]があり、同時適応については課題である。

これまでの研究では、適応度関数を構成するいくつかの重要なファクターが抽出されている。

#### 4 おわりに

本稿では、組織学習を個人の意思決定主体による意思決定に基づいてボトムアップ的に行われる過程とし

てとらえ、エージェントベースアプローチに基づいた組織学習を表現するための一般分配モデルを定式化した。

本稿で定式化した一般モデルでは、エージェントは2階層の集団を形成し、各エージェントは分配関数による配分を受ける。分配関数は、環境を表す係数である分配係数と、各エージェントの意思決定変数からなっている。エージェントは自己の変数と係数だけでなく、他のエージェントの変数と係数を認識し、内部モデルとして表現している。

本稿では、組織学習の基本的な4つ learning loop を定式化したモデル上で解釈し、モデルがそれらを表現していることを示した。また、double-loop learning における内部モデルの修正メカニズムと共有化のプロセスとして、遺伝的アルゴリズムによる進化的アプローチのステップを示した。

#### 5 参考文献

- [1] 安藤史江:組織学習と組織内地図, 白桃書房 (2001)
- [2] 出口弘:組織理論におけるエージェントベースアプローチ, 組織科学, 34, 2, 11/22(2000)
- [3] 高玉圭樹:組織学習の計算論的分析とオペレーション化, 組織科学, 34, 2, 36/45(2000)
- [4] 高橋真吾:システム理論からの組織論へのアプローチー組織学習の進化的システムモデルの枠組みー, 組織科学, 34, 2, 59/68(2000)
- [5] 高橋真吾:組織のダブルループ学習のための学習レベルを考慮した分配モデルによる基本的考察, 『社会組織のマルチエージェントシステム分析ー数理とシミュレーションからのアプローチ』システム工学部会・知能工学部会共催研究会資料 01PG0004, 121/126(2001)
- [6] W.Ross Ashby:Design for a Brain, John Wiley and Sons(1960)
- [7] R.Axelrod and M.D.Cohen:Harnessing Complexity, The Free Press (1999)
- [8] R.Axelrod, "The Complexity of Cooperation," Princeton University Press (1997).
- [9] C.Argyris and D.A.Schon:Organizational Learning II, Addison-Wesley (1996)

- [10]P.G.Bennet, "Hypergames: Developing a Model of Conflict," *Futures*, Vol.12, pp.489-507 (1980)
- [11]U.S.Putro,K.Kijima and S.Takahashi, "Simulation of Adaptation Process in Hypergame Situation by Genetic Algorithm," *SAMS*, (2001)
- [12]U.S.Putro,K.Kijima and S.Takahashi, "Simulation Approach to Learning Problem in Hypergame Situation by Genetic Algorithm," *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Vol.IV, pp.260-265 (1999)
- [13]G.Weiss (Ed.), *Multiagent Systems - A Modern Approach to Distributed Artificial Intelligence*, MIT Press, (1999)
- [14]S.Takahashi, "Evolutionary Approach to Three-person Hypergame Situation," *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Vol.IV, pp.254-259 (1999)
- [15]Dawid,H., *Adaptive Learning by Genetic Algorithms-Analytical Results and Applications to Economic Models-*, Springer (1999)(2nd ed.).
- [16]Espejo,R., Schuhmann,W., Schwaninger,M. and Bilello,U., *Organizational Transformation and Learning-A Cybernetic Approach to Management-*, Wiley (1996)
- [17]Hanappi,H., *Evolutionary Economics*, Avebury1994.
- [18]木嶋恭一編,マルチメディア社会システムの諸相,日科技連(1997)
- [19]Prietula,M.J.,Carley,K.M. and Gasser,L.(eds.), *Simulating Organizations*, MIT Press(1998)
- [20]Putro,U.S., Kijima,K. and Takahashi,S., "Adaptive Learning of Hypergame Situations Using a Genetic Algorithm," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, Vol.30, No.5, , pp.562-572 (2000)
- [21]高橋真吾, 中野文平, 荒瀬雅子「ハイパーゲーム的社会状況における認知の学習への遺伝的アルゴリズムの応用」*経営情報学会誌*, Vol.4, No1, ,pp.43-56 (1995)
- [22]高橋真吾 「ハイパーゲーム型社会状況における学習: コミュニケーションによる情報生成」 *進化経済学会論集第2集*, .314-322 (1998)
- [23]高橋真吾「組織の進化的学習のためのエージェント間インタラクションの基礎的考察」*経営情報学会秋季全国研究発表大会予稿集* pp.199-202 (1999)
- [24]Takahashi,S.,*Evolutionary Learning in Agent-based Modeling*,in H.S.Sarjoughian and F.E.Cellier(eds.),*Discrete Event Modeling and Simulation: A Tapestry of Systems and AI-based Theories and Methodologies, A Tribute to the 60th Birthday of Bernard P.Zeigler*,Springer(2001)
- [26]高橋真吾「エージェントベースシステムのための数理的システム理論の課題」*社会・経済システム学会誌*(2000)